GLOBAL OCEAN DATA ASSIMILATION: PROSPECTS & STRATEGIES

A USGODAE Workshop

April 23-25, 2001

University of Maryland Inn and Conference Center

November 2001

GLOBAL OCEAN DATA ASSIMILATION: PROSPECTS & STRATEGIES

A USGODAE Workshop

Sponsored by

NASA's Seasonal-to-Interannual Prediction Project

The Office of Naval Research

NOAA/Office of Global Programs

Report Edited by

Michele M. Rienecker, D. E. Harrison, W. Brechner Owens

November 2001

Workshop Steering Committee:
Eric Chassignet, RSMAS, University of Miami
D.E. Harrison, NOAA/Pacific Marine Environmental Laboratory
W. Brechner Owens, Dept. of Oceanography, Woods Hole Oceanographic Institution
Michele Rienecker, NASA/Goddard Space Flight Center
Paola Rizzoli, Massachusetts Institute of Technology

GLOBAL OCEAN DATA ASSIMILATION: PROSPECTS & STRATEGIES

A USGODAE Workshop

Executive Summary

Over the last decade ocean data assimilation software has reached a maturity to match that of atmospheric data assimilation. However we still are in the early stages of routine generation of global products, and, similar to the case for atmospheric products, the key to progress in this area is as much the supporting infrastructure as it is the assimilation tools themselves. This workshop on ocean assimilation was organized to help advance US activities in ocean data assimilation during the period leading up to the International GODAE Experiment (2003-2005). US ocean data assimilation activities include near-real-time mesoscale resolution efforts and basin-scale delayed-mode efforts.

The workshop report includes summaries of presentations, discussions, and the recommendations for priority activities that emerged from the discussions. From the workshop presentations and discussions, it was clear that some of the most important tasks to be undertaken are associated with the data streams – retrieval, development, and quality control (QC) of data sets, development of observational error characteristics – and delivery of the data for assimilation purposes. Although some of these activities can be accomplished through the cooperation between existing assimilation groups, most require an investment in and by the observational, data serving and assimilation communities to ensure that the progress is not made merely by serendipity. Thus, the GODAE patrons are requested to help facilitate funding of these activities through individual agency channels as well as through the National Ocean Partnership Program (NOPP).

The priority activities identified may be summarized as:

Data Set Development

- Organize standard, comprehensive quality controlled data holdings, particularly XBT and Argo for assimilation, and repeat hydrosections for validation and for observational error characterization
- Establish and document QC procedures.
- Develop value-added data sets, especially those to be used to assess the value to be added through model-data syntheses and those that ease the assimilation of the data.

Data and Product Serving

- The successful implementation and operation of the GODAE Server in Monterey is viewed as essential to achieving the goals of U.S. interests in GODAE. Priority needs to be given to stable configuration of, data set deployment on, and exercise of the server by assimilation groups.
- Increase usability of data and products on GODAE data servers for external users from the research community and by non-scientific groups.

Assimilation Product Development

- Analyze data archives to establish data error covariances, especially associated with representativeness errors for different models and different applications.
- Establish feedbacks between assimilation efforts and data set providers to help document data quality and effectiveness of quality control procedures.
- Develop assimilation tools that account for model biases.
- Conduct product intercomparison and validation through defined metrics.
- Establish forcing function errors.
- Establish better model and analysis error estimates.
- Investigate the impact of changes in the observing system on (1) mesoscale analysis and prediction, (2) seasonal to interannual analysis and prediction, and (3) on climate analyses and inferences from them.

Table of Contents

1. Workshop Goals, Format and Recommendations

The Global Ocean Data Assimilation Experiment (GODAE) is an international experiment with the goal of making routine ocean monitoring and prediction an operational activity similar to weather forecasting. For background, the scope of GODAE is presented briefly in Appendix A. U.S. contributions identified to date are summarized in Appendix B and include near-real-time mesoscale resolution efforts and basin-scale delayed-mode efforts. The U.S. GODAE Steering Team organized this workshop on ocean assimilation to help advance US activities in ocean data assimilation during the period leading up to the International GODAE Experiment (2003-2005).

The **goals of the workshop** were to:

- identify the factors that limit skill in ocean assimilation products for operations, forecasts and research
- identify improvement activities common to assimilation for mesoscale oceanography and climate forecasts and analyses;
- identify a few specific activities for concentrated community effort over the next two years;
- advance the development of metrics to assess progress in ocean data assimilation;
- foster GODAE pilot projects in the U.S.

To achieve these goals, the workshop took the form of presentations from invited speakers to summarize the state-of-the-art and identify issues in operational assimilation, in state estimation research, in data streams and in their use for assimilation purposes. These presentations set the context for follow-on plenary discussions. The workshop concluded with two break-out session discussions – one on metrics for GODAE products, and one on data issues for assimilation. The agenda is provided in Appendix C. Attendance was by invitation only, and the list of attendees is in Appendix D.

Summary recommendations

From the workshop presentations and discussions, it was clear that some of the most important tasks to be undertaken are associated with the data streams – retrieval and quality control (QC) of data sets, development of observational error characteristics – and delivery of the data for assimilation purposes. It was the opinion of our meteorological colleagues that the details – of data processing, of prescribed error covariances, and of model biases – can be more important than the assimilation technique itself. Hence, the commonalities between assimilation for operational oceanography and for climate analyses and prediction are stronger than the differences.

A primary need is the successful implementation and operation of the GODAE Server in Monterey. The successful implementation and operation of the GODAE Server in Monterey is viewed as essential to achieving the goals of U.S. interests in GODAE. Priority needs to be given to stable configuration of, data set deployment on, and exercise of the server by assimilation groups. A wide range of US GODAE needs assume the initial capabilities of the server will be in place in 2001. Additional development activities for the server are expected to continue throughout the operational phase of GODAE itself.

Other priority activities are summarized below. Although some of these activities can be accomplished through the cooperation between existing assimilation groups, most require an investment in and by the observational, data serving and assimilation communities to ensure that the progress is not made merely by serendipity.

Data retrieval

1. Foster activities to organize comprehensive data holdings:
   - undertake the development of "standard" QC'd XBT, Argo and drifter data sets;
   - establish and document QC procedures, including iterative QC;
   - establish metadata (flags, documentation) requirements for data.

Data processing

2. Develop processed data streams:
   - undertake development of assimilation-friendly datasets (e.g., altimetry, drifter data);
   - identify and prioritize value-added data sets (e.g., gridded, merged data streams such as sea surface temperature (SST) and sea surface height (SSH)), especially those to be used to assess the value to be added through model-data syntheses;

Data Serving

3. Foster activities on data handling and usability on GODAE data servers:
   - develop connections between servers for all data servers, including real-time and delayed mode data, to ensure interoperability;
   - establish connections to the NOPP Virtual Ocean Data Hub (VODHub) activities to provide GODAE input and ensure that developments fostered explicitly for GODAE are state-of-the-art and consistent with VODHub developments;
   - develop data management and serving capabilities, particularly for the GODAE server in Monterey, that promote data and product usage by external groups, not just by the scientific community.

Data and Observing System Issues for Assimilation

4. Foster activities to develop and improve data error covariances:
   - Representativeness error needs to be assessed by:
     - analyses of existing data, including a review of the literature,
     - incorporating new observations as part of process experiments to further this goal.

5. Establish feedbacks between assimilation efforts and data set providers:
   - Observation screening procedures are needed to discriminate between extremes and bad data.

- The data used by the assimilation needs to be monitored (e.g., as to whether the assimilation procedure rejected the data, document observation minus forecast values, etc.):
  - validation of data stream
  - feedback on data quality.

6. Foster activities to:
   - investigate the impact of changes in the observing system on (1) mesoscale analysis and prediction, (2) seasonal to interannual analysis and prediction, and (3) on climate analyses and inferences from them;
   - investigate the impact of data density on climatologies;
   - identify the need for new data.

Assimilation Product Development

7. Assimilation activities that are common between mesoscale and climate analyses were identified:
   - Metrics proposed: forecasts/hindcasts for mesoscale (24 hours to 30-days) and S-I (3 months with observed forcing).
   - Metrics should include information to calculate innovations (observation minus model forecast), observation minus analysis, with the assimilation statistics at observation locations, along with the analysis.
   - undertake cross-validation (by withholding data) as well as statistical consistency tests.
   - foster collaborative activities to establish better error estimates for model, data and analyses.

8. Infrastructure needs to be developed to promote use of GODAE products
   - facilitate assimilation product distribution through Monterey server;
   - identify prototype products and requirements for these to be made available through the server (formats, disk space, archive location, access issues).

The workshop report is organized as follows: a brief summary of presentations and plenary discussions is provided in Section 1. The reports from the two working groups formed as the final phase of the workshop are presented in Sections 2 and 3. Summaries from the invited speakers are presented in Appendix B.

2. Workshop Discussions

Assimilation: General
The presentations reviewed the state of the art of assimilation and of issues for the different applications within the ocean community. Assimilation for analysis needs to be cognizant of the interior sources/sinks introduced by traditional sequential filtering procedures used for state estimation. Smoothing procedures can be developed to maintain balances imposed by the model itself. The experience from the atmospheric assimilation provided a reality-check on the issues facing GODAE: the details are more important than complicated techniques. The important details include the error covariances (that are never known in practice, yet control the output from assimilation), data quality (especially the accommodation of data extremes during quality control) and consistency with the underlying model, and model biases. One of the unique problems of delayed-mode data assimilation is that the data holding is always changing because of data retrieval and correction.

Assimilation: Mesoscale
The discussions focussed towards identifying what currently limits assimilation for operational oceanography. These limitations were broken down into those associated with data, models, and assimilation:

- Data limitations:
  - inadequacy of operational observing systems;
  - inadequate operational infrastructure for processing and disseminating the data, including standardization of formats and QC flags;
  - observational errors include representativeness errors that are difficult to assess;
  - sampling biases;
  - knowledge of the mean sea surface height;
  - lack of good topographical data base.
- Model limitations:
  - evaluation of systematic errors in models;
  - sensitivity of models to accuracy and resolution of atmospheric forcing for both the analysis and forecasting cycle;
  - computer resources and software development for appropriate computer architectures.
- Data assimilation limitations:
  - inadequate knowledge of the covariance models that define both the analysis and forecasting problem;
  - extraction of meaningful information at the mesoscale resolution of the model from sparse observations, and how to take advantage of this information in an assimilation context;
  - subsurface projection of surface data;
  - lack of assimilation tools to account for model biases;

- cost-effectiveness of the assimilation method – need a method that addresses the major sources of uncertainty in the model and in the data.

One of the priority actions identified in these discussions was that of a national effort for the quality control of the historical XBT database, especially for the identification of outlier data. The XBT archives are not easy to manipulate, especially in the utilization of the QC flags. For mesoscale assimilation, the mean and standard deviation are not useful discriminators of data quality in regions of high variability. It would be useful to have a project to catalog XBT data according to local features, like fronts and eddies.

Assimilation: Climate scale

The discussions focused on the need for attention to data formats – the experience in meteorology is that standardization of formats has been key to the inclusion of different data sets – and on the need to assess trends inferred from assimilation products in the context of changes in the observing system. Again, limitations were broken down into those associated with data, models, and assimilation:

- Data limitations:
  - data formats need to retain all the meta information;
  - need to monitor data flow and data quality;
  - need to learn how to take advantage of new data types: drifters, gravity, color, salinity;
  - changes in the observing system can introduce spatial and temporal biases into analyses as well as impact climatologies.
- Model limitations:
  - evaluation of systematic errors in models;
  - representation of mixed layer and ventilation processes are sources of model bias;
  - computer resources are limited and impact the resolution used for climate analyses.
- Data assimilation limitations:
  - modeling the error sources;
  - accounting for biases;
  - choice of analysis variables;
  - appropriate measures of skill need to be identified, especially for studies designed to measure the impact of the observing system.

One of the priority actions identified in these discussions was that of a pilot project on data formats for both real-time and delayed mode data. Another was the need for continuous improvement and updates of data QC flags on the data server as data sets are improved over time. There are many issues that share commonality with mesoscale assimilation. Metrics should be tuned towards the end user, in this case primarily the seasonal-to decadal prediction applications and climate research. As for the mesoscale, forecast skill is a useful measure of assimilation product quality – even in an uncoupled model, a 3-month forecast skill can be useful. Climate data assimilation for research differs from that for the mesoscale in the importance of consistency between analyses, i.e., state estimation *per se* is not the only goal and internal sources and sinks are undesirable.

Data issues for assimilation
The discussions focussed on information available for assessing observational error covariances, including representativeness errors, and on the best way to use Lagrangian data. There are several examples of analyses of observational errors in the literature, but there has not been a significant, organized effort on this issue – the topic lacks excitement. In addition, there is a question about the adequacy of the observational data base to address these issues – variations in ship tracks, which can sometimes be displaced significantly, introduce uncertainties into statistics. Traditionally, observational errors are taken to be uncorrelated in space or time; however, for the climate problem, even the XBT error covariances are not diagonal because of representativeness errors. Unfortunately, estimating representation error is not as simple as estimating observation variance. One of the problems in addressing representation error is that it is model and resolution dependent.

Redundant data are needed to identify model bias, help with quality control, and estimate observational and model error covariances. Data sets that will be used as the basis for validation for assimilation products need to be identified. Two different metrics were discussed. The cost function minimized by the assimilation process can be used to test for statistical consistency. Independent (i.e., withheld) data can be used for cross-validation of the analysis. There needs to be some attempt at identifying suitable data to be withheld. Assimilation can be used as a QC check for observations. The development of this capability is a challenge for GODAE, but also promises to be one of its contributions to the observing system.

There is very little experience to date in dealing with Lagrangian data. One needs to deal with trajectories even in converting the data to an Eulerian representation.

Discussion of observational issues was continued in the Working Group on Data Characterization, as summarized below.

3. Working Group on Metrics

The metrics calculated for a particular analysis should take into account the underlying assumptions used to produce the analysis as well as the scientific purpose for which the analysis was intended. This will mean that the same set of metrics will not be produced for each GODAE analysis. Therefore, sufficient documentation must be available on GODAE servers so that a potential user can begin to assess the applicability of an analysis to a particular scientific question.

Mesoscale Forecast/Hindcasts

In ocean forecasting/hindcasting the timescale for predictive skill for some ocean features is closely linked to the time scale for atmospheric predictive skill while for others it is not. Quantities like sea surface temperature, mixed layer depth, Ekman surface currents, directly wind-driven coastal upwelling and shallow water currents fall in the first category. Mesoscale variability is primarily a consequence of flow instabilities and has a predictive time scale much longer than that for weather forecasts, a month or more. In addition, certain ocean internal waves such as equatorial waves, coastal Kelvin waves and Rossby waves, once initiated, can be predicted on time scales much longer than weather systems, for months even years. The oceanic response to El Niño/La Niña events is a striking example. Forecasts up to a month are appropriate for mesoscale variability, with atmospheric forcing reverting toward climatology beyond the time scale for weather forecast skill.

Metrics for global mesoscale-resolving forecasts are most useful if they include phenomenon specific and region specific measures, since the timescale for predictive skill is strongly dependent on these. In general, useful forecasts must be better than persistence and climatology as measured by statistics such as rms error, anomaly correlation, and the axis error of major upper ocean currents (all measured as a function of forecast duration). These measures are most useful if they are applied to fields that are strongly constrained by data such as sea surface height and sea surface temperature.

Seasonal-to-interannual initialization analyses

In addition to a validation of analyses against observations, the proposed metric for analyses used to initialize seasonal-to-interannual forecasts is a hindcast with a 3-month lead-time. The hindcasts should be driven by observed forcing fields to avoid the widely variable dependencies that would result from hindcasts performed by coupled ocean - atmosphere models. Both analyses and hindcasts should be provided in a form that enables their direct comparison.

Recommendation: A description/narrative of the limitations of the data assimilation procedure needs to be included together with information on how to calculate the innovations (i.e., observation minus forecast) and analysis error (approximated by observation minus analysis) as well as providing the analysis. Example of limitations: systematic heat import or water mass modification, thermocline diffusion.

In addition to forecasts/hindcasts, each data set needs to include the steps that were taken to validate them, i.e.,

- Comparison with independent data withheld from the data assimilation;
- Discussion of the processes present in the model: Are they physical? Are they artifacts of the data assimilation procedure?
- Consistency: Are the time series of water mass properties, volume, heat, and freshwater transports, etc, different from what is expected from observations or forward models alone?
- What is the value added by the assimilation procedure?

Statistical tests

In order to have a more objective intercomparison of model results, but also to help characterize the limitations of these comparisons, a set of standard measures should be selected. These will be problem specific (e.g., seasonal-to-interannual or mesoscale time scales) but it is nonetheless useful to have some common criteria for model validation. These measures mainly aim to evaluate how close these products are to reality. These measures will be:

- temporal mean of heat content in the upper 300 m

– rms (or standard deviation) of SSH compared to altimeter observations
– rms (or standard deviation) of SST compared to remote and in situ data.

In particular, for seasonal-to-interannual forecasts these will focus on the tropical Pacific, Atlantic, and Indian Oceans whereas for the mesoscale they will refer to the entire model domain.

Cross-validation: The use of Lagrangian observations for data assimilation is not yet feasible with present methodologies. However, because of the information carried by these observations, such as single and 2-particle statistics, they could be used for cross validation.

4. Working Group on Data Characterization

Data noise covariances need to be determined for all the assimilation efforts. At present many of these covariances are prescribed in an *ad hoc* manner. There are two main contributions to these covariances, instrument and representativeness errors. Both of these covariances contain off-diagonal components in that the errors for individual observations are often not independent. An example is the error for altimetric surface height observations where orbit error and uncertainties in the atmospheric corrections have scales of order 1000 km.

Most of the observationalists who are expert in the different data types have estimates of the instrument errors. However, their estimates are not always readily available, often buried in technical reports. We recommend that a survey of the literature be undertaken and a summary of these errors be maintained on the GODAE server.

The more difficult data error covariance involves the representativeness error. Besides depending on data type, it will also be sensitive to the model used for the assimilation since this error includes contributions from processes and scales that are not resolved by the model. Conceptually this covariance is related to an integral over the wavenumbers and frequencies of the signal spectra that are not resolved by the model but do contribute to a given measurement. For example, a sea surface height measurement will include contributions from mesoscale eddies that would not be included in a coarse grid climate model, so that the error covariance would include the mesoscale covariance. For a mesoscale model, a profile of temperature or salinity will include the influence of internal waves. Since the strength of the different physical processes will vary spatially and, probably, temporally this error covariance will likely be both non-homogeneous and non-stationary. There are a few existing data sets that are of sufficiently high resolution so that they can be used to estimate these error covariances. The high resolution VOS XBT sections are one example and the POLYMODE Local Dynamics Experiment is another. Although statistical analyses of these datasets have been done, they are not always published in the refereed literature. A survey of this literature should be completed and further analyses encouraged. Future process studies should also be encouraged so that the inhomogeneity of these statistics can be investigated. High-resolution models should also give some guidance on this issue. These future experiments would be similar to the POLYMODE LDE or Tropic Heat Experiment where the sampling was designed to be sufficient to resolve the signal covariances at scales smaller than the mesoscale. The experiments could also be used in regional very high resolution assimilation studies to diagnose the dynamics of the scales resolved by the increased sampling array.

Quality control (QC) of the data streams is an essential part of assimilation. This involves not just the preliminary step before the data is incorporated into an assimilation procedure, but will also involve comparisons of the products of the analyses with the data. Data centers are a necessary part of this process, providing a focus for the preliminary QC, providing a means to document the quality of the data, and providing a pathway for feedback to the data providers from the assimilators. To date GODAE has focused on the procedures needed to make the data "assimilation friendly." In a similar way, the assimilation products should include "data provider friendly" analyses to promote both further quality control of the data and monitoring of the observing system. At the least, the assimilation output should include the innovation (observation – forecast), and residual differences (observation – assimilation), and data covariance used in the assimilation. The data provider can then use these quantities to decide if the data is consistent with other nearby data and the model dynamics. For example, a bad profile from an Argo float or XBT would have a large normalized residual if it were inconsistent with nearby profiles and surface height measurements. The data provider and/or data center would use this information to set the quality flags associated with the data that are archived for either longer time scale forecasts or climate analyses. Comparisons of these misfits with other data points that are nearby in both space and time may enable the observationalists and assimilators to distinguish between legitimate extremes and bad data.

Surface drifters and sub-surface floats represent an under-utilized source of data for use in assimilation experiments. Although conceptually possible, the use of the trajectory data as model input will be difficult to implement and in the worst case be highly non-linear. Treatments of this Lagrangian data in an Eulerian frame are more straightforward. However, when the temporal resolution is low, for example the 10-day interval between positions for Argo floats, the problem becomes more troublesome. Temporal averages, for example bin-averages and dispersion statistics, should be more readily used as constraints. The surface drifter data, with a drogue at 15 m and noise due to windage, are problematic because they include ageostrophic effects that can be best estimated in a bulk layer sense. In the short term, it appears that the velocity data from floats and drifters will be best used for model validation. Research into the ways to use Lagrangian data in assimilation modeling should be pursued.

APPENDIX A

GODAE: Background to the Workshop

The Global Ocean Data Assimilation Experiment (GODAE) is an international experiment with the goal of making routine ocean monitoring and prediction an operational activity similar to weather forecasting. During the operational phase of the experiment, planned for 2003-2007, the integrated observing system established for routine operations will be used to used to provide regular, comprehensive descriptions of the ocean state by assimilating the data into state of the art models of the global ocean circulation in near real-time.

The scope of GODAE encompasses real-time, high-resolution, operational oceanography and near-real-time climate applications. The expected beneficiaries include climate analyses, seasonal-to-interannual forecasting, navy applications, marine safety, fisheries, the offshore industry and management of shelf/coastal areas. Observation networks, data access and distribution, models and estimation tools are all essential elements of GODAE. The U.S. contributions to GODAE, as identified to date, span the scope of GODAE and are summarized in Appendix E.

International plans for the experiment are described in the Strategic Plan (http://nsipp.gsfc.nasa.gov/usgodae/IGST/Strategic_Plan.pdf) and the implementation plans are currently under development. One of the underlying concepts of the strategic plan is the development of a GODAE Common – the infrastructure to support the routine generation of ocean products, both data-only and assimilation-based, the knowledge base accumulated through the conduct of pilot assimilation projects, and the assimilation products themselves.

During this development phase, prior to the conduct of the experiment, the U.S. activities are focused towards development of

- the necessary infrastructure, including reliable, routine data serving/distribution;
- coherent, organized data streams and data-only products for input to and assessment of the assimilation products;
- the shared knowledge base between assimilation groups and observationalists to optimize the utilization of the data and help improve the models and assimilation systems;
- the infrastructure for assimilation product delivery (including computational resources) that also promotes exploration, intercomparison and validation of the products;
- connections to user communities;
- metrics for evaluation of the products and the experiment.

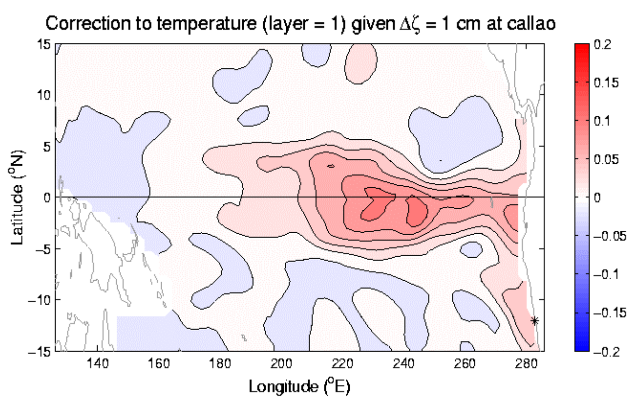
APPENDIX B

Summaries of Invited Presentations

Data Assimilation for the Large and Mesoscales

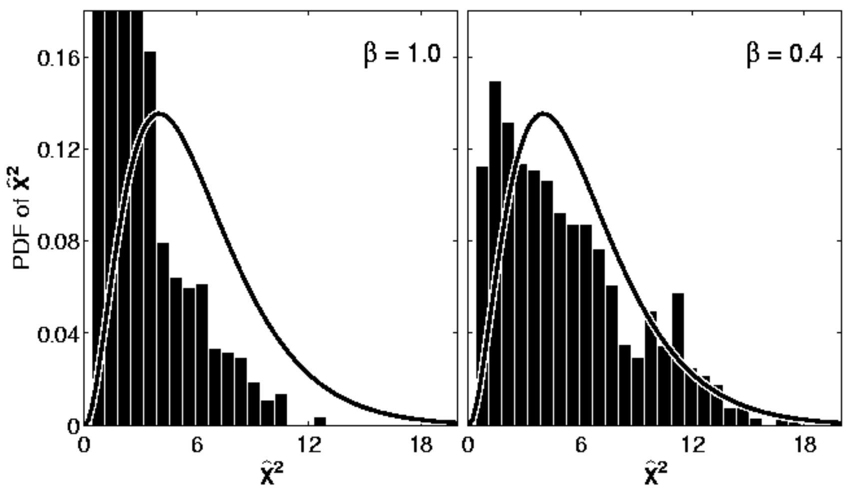
Robert N. Miller
College of Oceanic and Atmospheric Sciences
Oregon State University

- The process of data assimilation: what can we learn?
  - Most advanced data assimilation methods are formulated in terms of a positive definite quadratic cost function. In this framework, state space consists of a set of admissible functions of space and time.
  - This cost function can be seen to define an orthogonality relation, which in turn can be used to decompose state space into a subspace spanned by representers, one for each observation in space and time. The orthogonal complement of representer space consists entirely of unobservable states. Searching the unobservable subspace is the root cause of most convergence problems encountered in variational methods.
  - The support of a representer defines the region of space and time in which the corresponding observation influences the solution. The Kalman filter can be described within this framework. The columns of the Kalman gain are closely related to the representers.
  - Example: A column of the Kalman gain from a reduced state space Kalman filter applied to the Gent-Cane GCM.



This shows the extent of the influence on tropical Pacific SST of data from the tide gauge at Callao, shown here as *. Note that the maximum influence is along the equator, far from the tide gauge station itself. A model underestimate of sea level by 0.01m at Callao results in a maximum correction of about 0.1°C (Calculation performed and graph provided by R. Perez).

- Hypothesis Testing: distinguishing success from failure
  - The cost function $J(\mathbf{u})$ should be a random variable with χ^2 distribution when $\mathbf{u}$ is chosen to minimize J.
  - There is a corresponding test for filtering algorithms. A χ^2 variable can be manufactured from the innovation vector and the prior error covariance estimates.
  - This is one basic test for consistency of the prior error estimates.
  - Other self-consistency tests are available.



Histograms of the χ^2 variable for different choices of parameters in prior model error covariance in an OI scheme for the Oregon coast. Curves are χ^2 pdf's. β is the relative magnitude of forecast and observation error variances. From Oke et al., 2002.

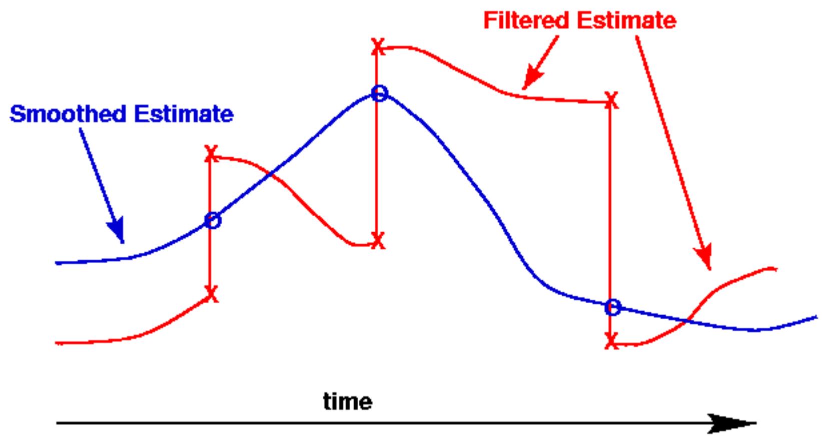
- Nonlinearity
  - So far our linearized methods have served us well. Are there clouds on the horizon? Example: multiple paths of the Kuroshio. Tide gauge data can detect the formation and decay of the meander in the Kuroshio off Japan. Are these data sufficient to constrain a data assimilation system to produce the correct evolution of the Kuroshio path over time scales of decades?

Assimilation for Climate Analyses

Ichiro Fukumori
NASA/JPL

Goals of assimilation:
"Operational" → Estimation of the state
"Research" → Explanation of the state

- State estimation does not require as stringent and as accurate an error estimate as needed for explaining the state, but it helps.
- Filtered state estimates do not provide physically consistent explanations of the state.
- The factors most limiting to GODAE at present are:
  1. Incomplete modeling and treatment of
     a. process noise
     b. representation error
  2. Resources
     a. computational
     b. human.



Time-integrated filtered dynamics: Evolution of red filtered curves

Time-integrated smoothed dynamics: Evolution of blue smoothed curves.

Model: $\mathbf{x}(t+1) = F(\mathbf{x}(t), \mathbf{w}(t)) + \mathbf{q}(t)$

Filtered State: $\mathbf{x}_f(t)$
Filtered Dynamics: $F(\mathbf{x}_f(t), \mathbf{w}(t)) - \mathbf{x}_f(t)$
Smoothed Dynamics: $F_s(\mathbf{x}_s(t), \mathbf{w}_s(t)) - \mathbf{x}_s(t)$

Filtered estimate: $\mathbf{x}(t) = \mathbf{x}(t,-) + \mathbf{K}[\mathbf{y}(t) - \mathbf{E}\mathbf{x}(t,-)]$
Kalman filter: $\mathbf{K}(t) = \mathbf{P}(t,-)\,\mathbf{E}^T\,[\,\mathbf{E}\mathbf{P}(t,-)\,\mathbf{E}^T + \mathbf{R}]^{-1}$
$\mathbf{P}(t)\,\mathbf{E}^T\,\mathbf{R}^{-1}$

Smoothed estimate: $\mathbf{x}(t,+) = \mathbf{x}(t) + \mathbf{S}[\mathbf{x}(t+1,+) - \mathbf{x}(t+1,-)]$
Smoother: $\mathbf{S}(t) = \mathbf{P}(t)\,\mathbf{A}^T\,\mathbf{P}(t+1,-)^{-1}$

Model error covariance matrices **P** ought to satisfy certain physical conditions: $\mathbf{HPH}^T=0$;
e.g., no errors in total heat content for model errors due to winds and finite differences.

It's easier to model process noise **Q** instead of **P**.

$$J = \sum_t (\mathbf{y} - \mathbf{Ex})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{Ex}) + \sum_t (\mathbf{x}(t+1) - F(\mathbf{x}(t)))^T \mathbf{Q}^{-1} (\mathbf{x}(t+1) - F(\mathbf{x}(t)))$$

The priors **Q** and **R** together define the assimilation problem.

Model constrain error (process noise) **Q**:

$$\mathbf{Q} = \langle \mathbf{qq}^T \rangle \;;\quad \mathbf{q} \equiv \bar{\mathbf{x}}(t+1) - F(\bar{\mathbf{x}}(t), \mathbf{w}(t))$$

e.g.,

- Errors in forcing, boundary conditions, model parameters
- Numerical inaccuracies due to finite-differences or spectral approximations
- Effects of unresolved processes on **x**, e.g., eddy contribution to large-scale heat/fresh water/etc. budget, absence of deep transport through fracture zones.

Data constraint error **R**:

$$\mathbf{R} = \langle \mathbf{rr}^T \rangle \;;\quad \mathbf{r}(t) \equiv e + (\bar{\mathbf{y}}(t) - \mathbf{E}\bar{\mathbf{x}}(t))$$

e : Measurement error, e.g., instrument accuracy, ship position error, bio-fouling, instrument correction error.

$(\bar{\mathbf{y}}(t) - \mathbf{E}\bar{\mathbf{x}}(t))$: Representation error, e.g.,

- model state represents an averge around a grid point; coarse resolution models do not resolve eddies
- QG models do not simulate diabatic processes
- reduced gravity models do not account for barotropic processes
- Argo float samples a meddy

Issues in **Q** and **R** prescription:

- Modeling error sources other than surface forcing, e.g., mixing parameterization, time-mean stratification, water-mass formation, unresolved topography, grid resolution.
- Dealing with time-correlated errors
- Dealing with biases
- Non-white representation error
- Designing efficient state reductions to approximate errors.

What can be learned from operational atmospheric data assimilation experience

John C. Derber
National Centers for Environmental Prediction
Environmental Modeling Center

Details often more important than basic technique

- Data handling infrastructure (quality control, data availability and use, formats, monitoring)
- Choice of analysis variables
- Specification of error covariances
- Quality of forecast model (and boundary conditions)
- Bias correction of forecasts and observations
- Feedback between various components of system

Data handling infrastructure

- Top reasons for not using data
  - Lack of useful information
  - Unreliable
  - Unstable
  - Insufficient quality control
  - Insufficient forward model
  - Unique format (too much work to get in)
  - Data unavailable (in real time)
- Quality control
  - Instrument specific
  - In conjunction with analysis (between different types of data)
  - A few bad data points can do more damage than many good data points
  - Correlated errors especially difficult to Q.C. (All remotely sensed data has correlated errors)
- Data availability and use
  - If data is not available in time it will not be used
  - Real time monitoring and use of data most efficient
  - Forward model (and adjoint) are necessary for each type of data – quality of this model very important
  - All quantities needed for forward model for observation should be in data file

- Use data as close as possible to source (especially satellite data)
- Observational and representativeness error (including forward model)

- Formats
  - No one wants to deal with many data sets in different formats
  - Should allow incorporation of all necessary (and potentially necessary observational information, diagnostics, and events
  - Once established difficult to change

- Monitoring of data
  - Essential for diagnosing problems with observations, model, forcing and data assimilation system
  - Inclusion of new data sources in monitoring necessary before inclusion in data assimilation system
  - Creates necessary statistics for inferring error statistics
  - There is never enough monitoring

Choice of analysis variables

- Analysis variables do not have to be same as model or observational variables
- Analysis variables need to include all variables necessary to properly simulate observations
- Analysis variables can be chosen to simplify structure of background error covariance and to include balance

Specification of background error covariance

- 2 problems
  - Computational ability to properly specify error covariance
  - Estimation of error covariances
  - Currently considered most important problem in atmospheric data assimilation

Quality of forecast model (and boundary conditions)

- Quality of forecast important for determining usefulness of data
- Changes in forecast → changes in statistics
- Improvements in forecast model results in larger than expected improvements in forecasts because of improvements in data assimilation

Bias correction of forecasts and observations

- Most data assimilation theory assumes unbiased observations and model – not true
- I think this is the most important overlooked aspect of data assimilation.
- Removal of bias from satellite data necessary prior to use

- Biases in background fields particularly large (compared to signal) in ocean. Should be removed prior to analysis

Feedback between various components of system

- All components of data assimilation system interact strongly with all other components
- The weakest link determines the quality of the results
- Data impacts can be overestimated with poor model, inadequate assimilation system or with incomplete use of data base
- Development of appropriate measures of skill

Final comments

- Poor design of data handling system results in many delays and significant additional work
- Data or models cannot be used blindly
- The development of a quality data assimilation system requires all components to be high quality
- Data assimilation monitoring and feedback to assimilators and modelers is essential
- Much of improvement to assimilation system incremental
- Change management

Practical Aspects of Error Covariance Modeling in Atmospheric Data Assimilation

Dick Dee
Data Assimilation Office
NASA/GSFC

The Analysis Problem

$$J(x) = (x^b - x)^T P^{-1}(x^b - x) + (y^o - H(x))^T R^{-1}(y^o - H(x))$$

- Linearized solution:

$$\begin{aligned} x &= \arg\min J(x) \\ &= x^b + PH^T[H\,PH^T + R]^{-1}[y^o - H(x^b)] \end{aligned}$$

- The observation operator H may involve extrapolation in time (4DVAR)
- The nonlinear case involves successive linearizations
- P and R define both the problem and the solution

In Theory: Optimal Estimation

- P and R should reflect the accuracy of the information being analyzed. In the linear case,

  if $P = <e^b (e^b)^T>$, $R = <e^o(e^o)^T>$
  and $<e^b> = 0$, $<e^o> = 0$, $<e^o(e^b)^T>$,
  then x^a is the estimate with minimal error covariance:

$$P^a = P - PH^T[H\,PH^T + R]^{-1}HP.$$

- error reduction takes place only in the observable subspace of the range of P
  - in theory this is precisely what you want, but …
  - in practice it implies a major constraint imposed by the covariance model

Error Evolution and Model Errors

- Cycling in time (k): $P(k) = <e^f(k)(\, e^f(k))^T>$
- Error contributions:

$$e^f(k) = M(e^a(k\text{-}1)) + e^m(k)$$

  - model-propagated initial error
  - model-propagated initial error
  - accumulated model error

- If random model error is large then it must be accounted for, because the solution of the approximate covariance evolution equations

$$\begin{aligned} P(k) &= M\, P^a\,(k\text{-}1)\, M^T + Q(k) \\ P^a\,(k) &= [I - K(k)\, H(k)]\, P(k) \end{aligned}$$

  is driven by Q(k) wherever data are sparse (Cohn 1993)

- Modeling/estimation of analysis error covariances is very difficult
- Hard to see how one can ignore model error in ocean data assimilation
- A large part of model error is systematic rather than random

Covariance Models Used in Current Practice

- Probability theory serves as a guide, but cannot be taken too seriously:
  - many assumptions needed for the theory are not realistic
  - we have very little information about actual error characteristics
  - despite appearances, the crux of the problem is not computational
- Approach used in operational atmospheric data assimilation centers:
  - explicit parameterization of multivariate and spatial correlations
  - quasi-stationary and quasi-separable
  - background error covariances are not flow-dependent
  - observation error covariances are exceedingly simple
  - variances may have some temporal (seasonal) evolution
  - spectral formulations: model fitting by means of the "NMC Method" or variant
  - physical-space formulations: covariance parameter estimation using observations

Using Proxy Data to Estimate Covariance Parameters

- The most popular method in use at NWP centers:
  - use (24-h forecast minus verifying analysis) as proxy for forecast error (The "NMC method," Parrish and Derber 1992)
  - fit parameterized covariance model to time-mean covariance of proxy
  - ECMWF now uses ensemble of analysis experiments with perturbed observations and stochastic physics
- Advantages:
  - relatively easy to obtain global, multivariate covariance specifications
  - friendly to the model
- Drawbacks:
  - relation with actual errors is questionable, especially in data-sparse areas
  - impossible to get small spatial scales
  - seasonal time scales, but no flow-dependence
  - arguably, what you put in is what you get out

Using Observations to Estimate Covariance Parameters

- Requirements:
  - must account for observation errors (there is no truth)

- must be able to parameterize the problem (not enough observations)
- must choose an appropriate analysis variable (well-defined statistics)
- must be careful about quality control and bias

- Techniques:
  - binning of station residuals (Gandin 1963)
  - Generalized Cross-Validation (Wahba 1980)
  - maximum-likelihood estimation (Dee 1995)
  - Bayesian approach (Purser & Parrish 2000)

- Problems
  - identifiability (separation of observation and background errors)
  - must have a good covariance model (especially if flow-dependent)
  - sampling error (must do some kind of averaging)
  - numerical optimization (nonlinear cost function with singularities)

Some Current Directions in Covariance Modeling being pursued in global atmospheric data assimilation

- Physical-space covariance parameterization (Cohn 1993; Riishøjgaard 1998):
  - intuitive flow-dependent modification of isotropic correlations
  - dynamical model for variance evolution
  - relatively difficult to implement, relies strongly on the background field

- Ensemble-based covariance approximations (Evensen 1994)
  - the choice of the ensemble is controversial
  - relatively simple to implement, handles nonlinear error evolution
  - must take care to avoid artificial remote correlations
  - covariances are not full rank

- Several groups are trying to combine these approaches (Mitchell and Houtekamer 2000; Heemink et al. 2002)

Choice of Analysis Variable

- Choose an analysis variable which facilitates covariance modeling
- Example: moisture (atmospheric water vapor)

Model Bias

- Lots of evidence that actual errors have large systematic components
- In combination with a changing observing system, model bias will induce spurious climate signals
- Many examples of manifestations in atmospheric reanalysis data sets

Adaptive Estimation and Correction of Model Bias

- It is not difficult to estimate and correct slowly varying errors in real time (Dee and da Silva 1998)
- Sequential moisture bias correction is now operational in the TERRA data assimilation system at the DAO (Dee and Todling 2000)
- Complete multivariate bias correction has been implemented
- A simplified version of the bias correction scheme is essentially cost-free
- It is possible to generalize to other discernible signals in the model error, such as errors in the diurnal cycle (Radakovich et al., 2001)

Some Methods for Assessing the Error Models

- Ability to predict future data:
  - observed-minus-forecast statistics
  - forecast skill
- Consistency of the assumptions with observations:
  - chi-squared statistics (necessary but not sufficient)
  - monitoring of statistical quality control
- Time evolution of the errors:
  - station data
  - model bias estimates
- Total response of the end-to-end system:
  - diagnostic quantities (precipitation, OLR, …)
  - climate parameters (strength of the Hadley cell, …)
- Sensitivity studies:
  - impact of observations
  - adjoint of the data assimilation system

Summary

- Covariance models define the analysis problem as well as its solution
  - beware of constraints imposed by the error models
  - avoid circular arguments
- In operational data assimilation systems, covariance models are based on practical and physical considerations rather than on probability theory:
  - imposition of balance requirements
  - computational efficiency
  - tuning to bottom-line performance

- Many aspects of error modeling are still unsolved:
  - realistic flow-dependent covariance models
  - time behavior of the errors
  - observation errors (including representativeness)
- In some cases simplification leads to improved performance, and to a better starting point for further development
- We don't know yet how to extract meaningful information at the resolution of the model from the observations

Global Coupled Ocean/Atmosphere Analysis/Prediction System

James A. Cummings
NRL/Monterey

The objective of our work is to develop, test and validate a global coupled data assimilation system comprised of atmosphere and ocean components. Each component will contain programs to perform data quality control, data analysis, initialization and numerical forecasts. Our approach is to build the coupled system using a combination of existing and newly developed components and a generalized flux coupler to allow for the exchange of relevant parameters across the air-ocean interface. For the atmosphere, the Navy Global Atmospheric Prediction System (NOGAPS) will be used for the atmospheric forecast component of the system. Atmospheric data assimilation is performed using a three dimensional variational (3DVAR) analysis. For the ocean, the Parallel Ocean Program (POP) model will be used for the ocean forecast component of the system. Oceanographic data assimilation will be performed using a three dimensional multivariate optimum interpolation (3DMVOI) analysis. The flux coupler controls the exchange of heat, momentum and moisture fluxes across the air-sea interface between the ocean and atmosphere models. The flux coupler has been written in a modular form to allow a variety of flux correction algorithms to be tested within the framework of the coupled system. In operations, it is expected that the global coupled system will provide improved capabilities to describe the atmosphere and ocean at the analysis and forecast times, and to provide high resolution initial and boundary data for the atmosphere and ocean mesoscale system. The global coupled system is being designed for real-time operational use at Fleet Numerical Meteorology Oceanography Center.

Operational atmospheric data assimilation is a mature technology practiced at several national weather centers around the world. Ocean data assimilation, on the other hand, is far less advanced than for the atmosphere, mainly because of the lack of observational data and the large computational requirements of ocean circulation models. Recent advances in observing systems and computer technology are finally removing these obstacles, but there are still many challenges that remain. The issues confronting successful implementation of the coupled data assimilation and forecast system being developed by NRL can be grouped into several categories: models, observations, and analysis techniques. As will be shown, many of these issues are interrelated.

Model Issues

During the forecast period the NOGAPS and POP models will tend towards their preferred model states. This issue is the problem of model climate drift. We hope to minimize climate drift of the models by implementing a coupled system but, due to operational constraints, we will not be able to execute the real-time system in a tightly coupled mode. For the initial operational capability we will only be able to provide NOGAPS with a time dependent SST lower boundary produced by the last forecast run of POP, which was forced by the last forecast run of NOGAPS, and so on. The atmosphere and ocean data assimilation update cycles will constrain the forecast models where there are observations, but in the long term absence of observations the atmosphere and ocean models will evolve unconstrained. At the present time we have very little quantitative knowledge of NOGAPS and POP model errors and biases when the models are executed in a data assimilative coupled system. If biases are detected, the model should be corrected and the biases removed, hopefully by tracking the biases back to identifiable causes. For example, we have recently implemented in NOGAPS a new cloud cover prediction scheme for cumulus and stratocumulus. The new scheme is able to reproduce the distribution of subtropical boundary layer clouds in a much more realistic way. As a consequence, the short-wave surface heat flux is estimated better in these areas, alleviating to a large extent the large positive SST biases that have occurred in the past. The process of quantifying model errors and biases at depth is more difficult, however, because of the scarcity of subsurface data in the ocean and the sampling limitations of the operational data stream. Nevertheless, identifying and removing model errors and model biases is an on-going effort in our implementation of the coupled model system. We look to new observing systems such as ARGO to help in this regard.

Observation Issues

Initially, we need to improve our use of the current suite of observations. We need to understand which operational observing systems are the most useful to the coupled system. This effort will require improved knowledge of the error characteristics of the data and what constitutes noise as opposed to real small-scale structure supported by the observations. For some observing systems we need to reduce the volume of the data before attempting to assimilate the data into the model. The preprocessing of these data will require a good understanding of the fine-scale phenomena we are attempting to forecast.

There needs to be coordinated monitoring, quality control and bias removal of the operational observing systems. Quality control is state dependent and adaptive methods need to be developed to distinguish between truly erroneous observations and valid observations spuriously rejected due to an extreme event. The error statistics used in the quality control evolve with time, but reflect average, rather than extreme, conditions in the ocean. When large, rapid changes are taking place that are not reflected in the ocean forecast state, we need to adjust the error tolerances in the quality control process to prevent rejecting good, and likely very important, observations.

Temperature is routinely measured in the ocean but for multivariate methods, where density and geopotential are additional analysis variables, companion salinity observations are needed for every temperature observation. A variety of strategies exist for solving this problem, from climatological TS relationships to using the model salinity field, all with varying degrees of success. The ideal solution, of course, is to increase the number of real-time salinity observations in the world's oceans.

Finally, we need to prepare for future observation types. Some of these types will have large data volumes and there will be network design issues to ensure proper distribution of the data in real-time. In addition, some of these new data types will not be direct observations of analysis variables thereby requiring forward models to convert from one data type to another for the data assimilation.

Analysis Issues

In the global coupled system the primary purpose of the data assimilation is to provide initial conditions for the forecast models. The assimilation is performed using a sequential incremental update cycle and the analysis is based on data observed up to and including the analysis time. The measure of success is the accuracy of the forecasts issued from the analyses. However, one of the issues we wish to address is the assessment of analysis quality independent of forecast quality. We need to develop and implement objective, a posteriori, methods of doing this.

The relative importance of one data assimilation scheme versus another for practical applications has not been addressed. The ocean data assimilation method we will be using in the coupled system is a static method that does not take error dynamics into account. However, we clearly need situation and flow dependent analysis techniques. Unfortunately, methods that take error dynamics into account are much more computationally expensive than the OI based methods we will be using. Recall that the global coupled system must be executed in near real-time under strict wall-clock time limitations set by the operational center. Many approximations have been made to the more advanced methods to render them more similar to the static methods in terms of computational expense, but the approximations are still a lot more expensive. Thus, it is an open question as to whether the benefits to be gained by these approximations are worth the additional cost in the large-scale system we are developing. We will look to the experiences of the other data assimilation groups in GODAE for guidance on this issue.

The most critical inputs to the data assimilation process are the observational error and forecast error covariances. The data assimilation methods we will be using in the coupled system require this information to be specified a priori. Of the two components of the error covariances the correlations, which control the way information is spread, may be considered more important than the variances. The analysis is more sensitive to the specification of the background error correlations in data sparse areas than in data dense areas, yet the standard innovation correlation method that is used to compute correlation length scales and observation and forecast errors requires a dense data network. This requirement is contradictory. We need to continue to refine our error estimates for the observations, the models and the initial conditions. The use of wrong a priori statistics will lead to erroneous results. We need to investigate hybrid algorithms that break the assimilation into two estimation loops, one for the state variables and one for the error parameters. If the forecast error covariances evolve on time scales longer than the time scales of the flow, then the two estimation loops could be executed at different times, thereby rendering such a system feasible for real-time applications.

- Finally, we need to investigate the need for an initialization step in the ocean data assimilation to prevent “shocking” the model when starting from the analyzed initial conditions. This effort, in and of itself, is not a trivial problem to solve.

Results from the operational ocean analyses at NCEP/CMB

Dave Behringer
National Centers for Environmental Prediction
Environmental Modeling Center

- Based on an ocean general circulation model (MOM)
  - The operational version has been a Pacific Ocean configuration of MOM1
  - A current experimental version is a global configuration of MOM3
- The assimilation system is a 3D variational scheme (Derber and Rosati, 1989)
  - Original system assimilated only temperature observations
  - Modified in the mid-1990's to assimilate the SSH variability observed from TOPEX
  - Modified again more recently to assimilate salinity as well as temperature and SSH
  - Latest version has been re-coded to run on distributed processes using MPI

Assimilating SSH

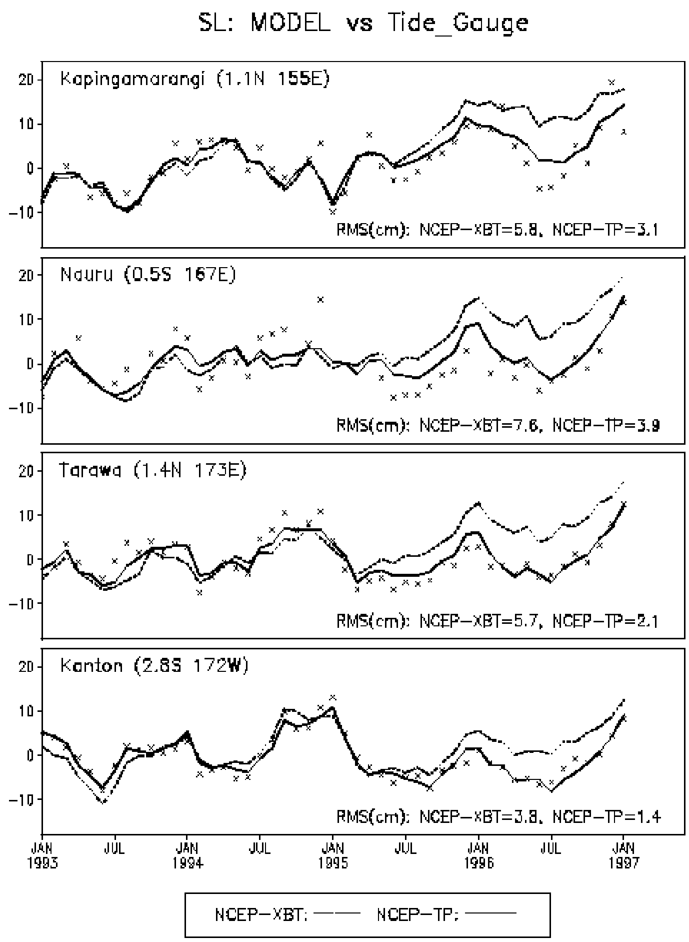
- Only the variable part of the observed SSH is used
- In the initial configuration only temperature was corrected
- The specification of the background error covariance completely determines the vertical distribution of temperature correction (in the present model of the error covariance the local error variance is assumed to be proportional to the local vertical gradient of the background temperature, which has the effect of concentrating corrections in the thermocline)
- The assimilation of SSH can potentially improve the model representation of temperature and SSH in time and space where T(z) observations are sparse
- However,
  - in a univariate system, errors in the mass field that may be due to salinity errors can be wrongly folded into the temperature correction
  - inserting only temperature corrections into an ocean model will lead to an important degradation of the model salinity

Assimilating Salinity

- By assimilating salinity, the model salinity can be corrected and maintained and not destroyed as it is in a system that corrects only temperature
- In a bivariate system that corrects both salinity and temperature, SSH can be assimilated and corrections can be apportioned between temperature and salinity
- However, in the absence of S(z) observations, some means must be available for estimating salinity profiles (in the present system this is done by a statistical method that preserves some of the temporal variability in the local T-S relationship).

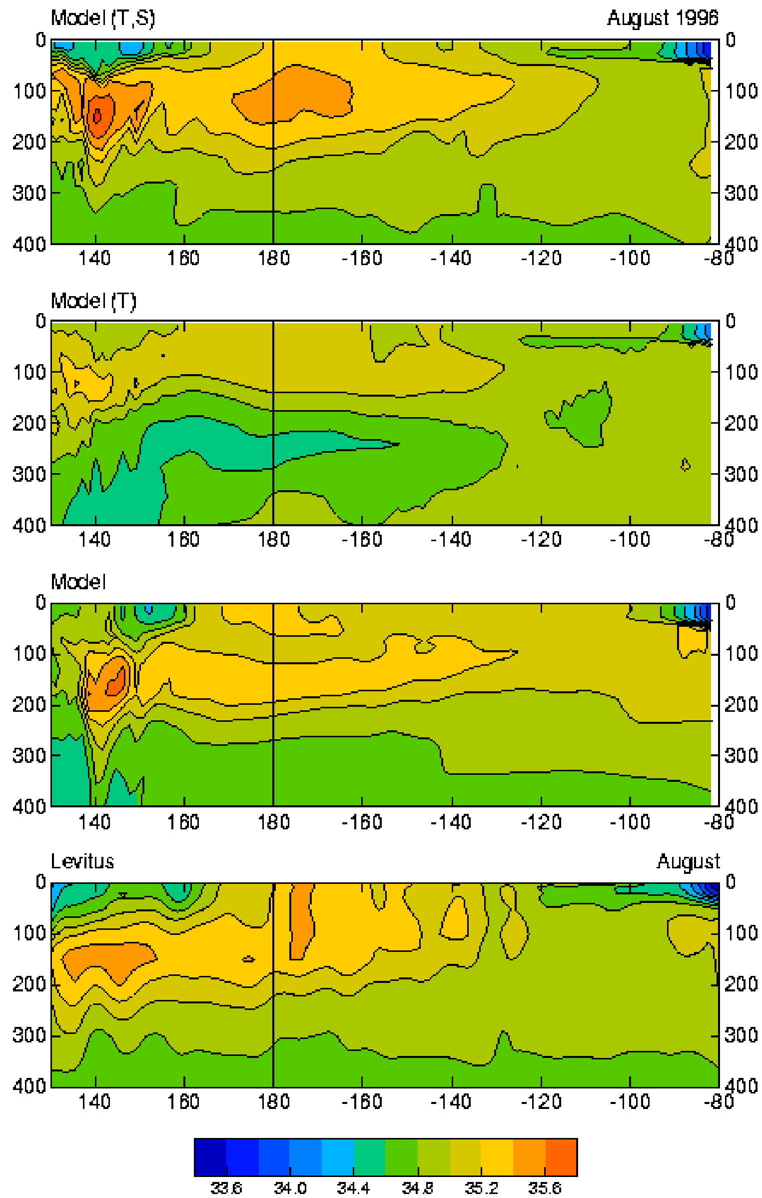
Future plans for the CMB Ocean Data Assimilation (ODA) System

- Explicit handling of model biases
- Improved background and observational error covariances
- Development of a 4D variational ODA based on MOM4
- Development of an ensemble filter ODA

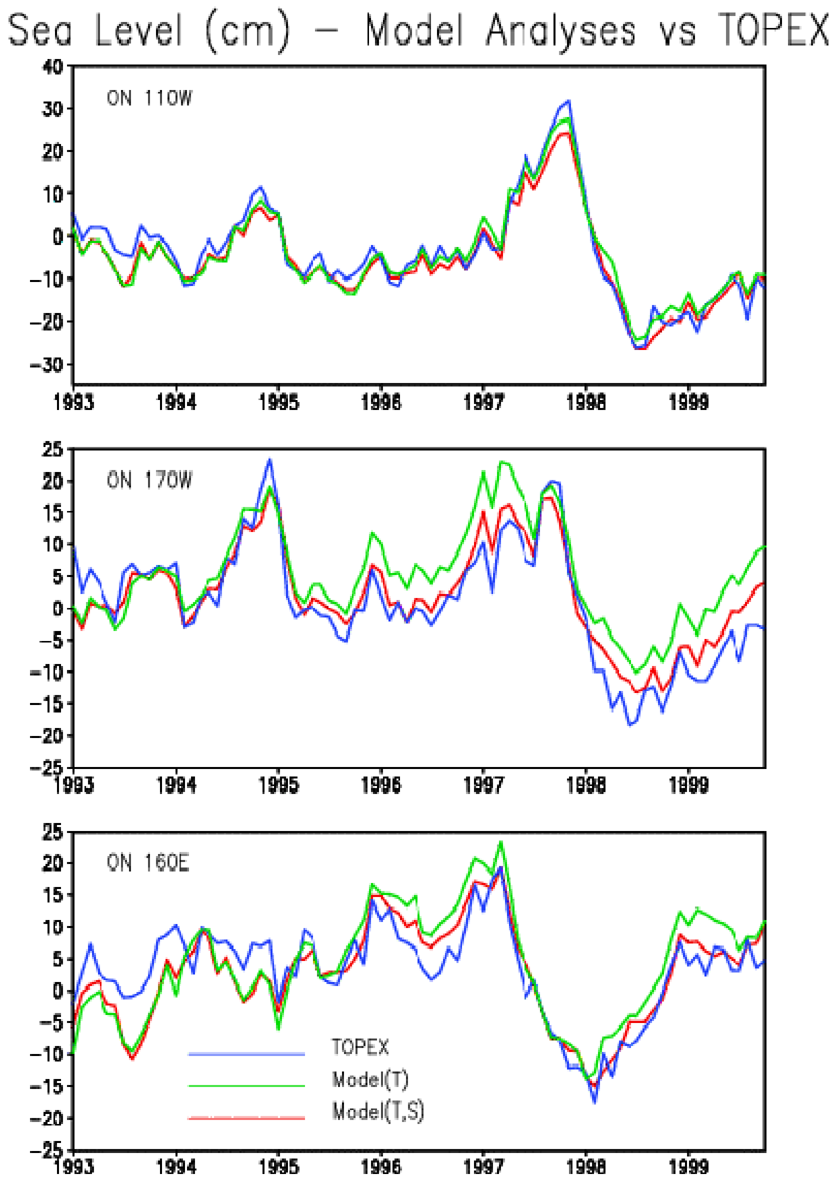


A comparison of sea level from two model analyses with independent tide gauge data. One analysis assimilates only temperature (NCEP-XBT), and the other assimilates temperature and altimetry, (NCEP-TP). It appears to suggest two things. First, there are times when temperature data alone are sufficient to constrain the model solution to reproduce observed sea level in the western tropical Pacific and times when it is not. Second, the additional constraint provided by altimetry will ensure that the model reproduces sea level at all times, but with the implication that, at times, it is at the cost of a larger and systematic error in temperature.

Salinity ($^{o}/_{oo}$) Equator



A comparison of analyses of equatorial sections of salinity. The first three are from model analyses that assimilate both temperature and salinity (Model(T,S)), temperature alone (Model(T)) and no data at all (Model) and the last is from Levitus. They illustrate the degradation of the model salinity field in the tropics that occurs when only temperature is assimilated. It demonstrates the importance of assimilating salinity as well as temperature if water mass characteristics are to be conserved.



An illustration of the improvement in the model representation of sea level for the case where both temperature and salinity are assimilated as compared to the case where only temperature is assimilated. The model runs are compared to TOPEX altimetry, which is independent data for these two runs.

Uncertainties of in situ Eulerian Observations

Paul Robbins
Physical Oceanography Research Division
Scripps Institution of Oceanography

- Examples of measurement and representation errors
  - Strategies for determining data covariances from observations
  - Building <$\mathbf{rr}^{T}$> has greater data requirements than assimilation step

- Representation errors from
  - time series [e.g., Fillenbaum et al., 1997]
  - repeat sections [e.g., Roemmich and Gilson, 2001]

- Measurement errors from
  - multiple measurement methods [e.g., Wijffels, Toole and Davis (2001) analysis of rms differences between altimetry and hydrography along the P6 WOCE section: rms difference was ~8cm. Of this 4 cm is estimated due to internal wave aliasing → 6.8 cm for altimetric error.]
  - Analysis

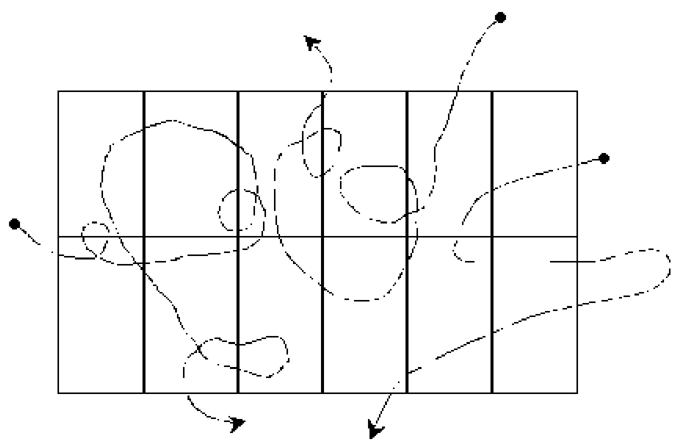
- Even for XBT data, representation error is not diagonal for climate data assimilation. We need more data to estimate these covariances.

- Mesoscale representation error is mainly from internal oscillations.

Aspects of Oceanic Lagrangian Data

Joe LaCasce
Physical Oceanography Division
Woods Hole Oceanographic Institute

Single particle statistics



Measures

$$\mathbf{U}(x, y, z), \qquad \sigma^2(x, y, z) \equiv \langle(\mathbf{u} - \mathbf{U})\rangle^2$$

$$\kappa_1 \equiv \frac{1}{2}\frac{d}{dt}\langle X^2(t)\rangle = \frac{1}{2N}\sum_i [(x_i(t) - x_i(0))^2 + (y_i(t) - y_i(0))^2]$$

Davis (1991), Owens (1991), Swenson and Niiler (1996)

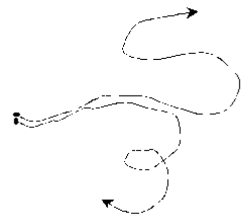
Pros:

- Absolute surface and deep velocities
- Extensive regional coverage
- Atlantic, Pacific, Indian, Southern Oceans
- ARGO (note: hydrography)

Caveats:

- Non-uniform sampling (errors)
- Isobaric vs isopycnal sampling
- Continuous and segmented temporal sampling

Two particle statistics



Measures

$$k_2 \equiv \frac{1}{2}\frac{d}{dt}\langle D^2 \rangle = \frac{1}{4N}\sum_{i \neq j}[(\mathrm{x_i} - \mathrm{x_j})^2 + (\mathrm{y_i} - \mathrm{y_j})^2]$$

$$v_{rel}^{\;2} \equiv \langle (\tfrac{dD}{dt})^2 \rangle = \frac{1}{2N}\sum_{i \neq j}[(\mathrm{u_i} - \mathrm{u_j})^2 + (\mathrm{v_i} - \mathrm{v_j})^2]$$

Davis (1985), LaCasce and Bower (2000)

Pros:

- Direct measure of error growth
- Scale-dependent separation characteristics

Caveats:

- Pair deployments rare
- Slow statistical convergence

Other statistics

1) Single Particle dispersion relative to topography

Subsurface North Atlantic

- Sensitivity at all depths
- Nearly isotropic dispersion

Subsurface North Pacific

- Little sensitivity
- Strongly zonally anisotropic dispersion

2) Velocity PDFs

Subsurface North Atlantic

- Weakly non-Gaussian
- Similar PDFs in unforced 2D turbulence

Conclusions

Lagrangian data exhibit strong variability and rapid error growth

- Turbulent ocean (?); regional variations; statistical characterizations sensible

Relation to models:

- validation (statistical comparisons); data assimilation; understanding the data (simplified models)

Observation representativeness and sampling issues

Bruce Cornuelle
Scripps Institution of Oceanography

The goal of the presentation was to help frame a discussion of issues on the observational side of any data assimilation system. The issues include: observation errors (especially correlated ones), quality control limitations (especially space-time coverage), model representation errors, observing system design. For specificity, points were illustrated using the High-Resolution VOS XBT (HRXBT) network, and data analysis by Dean Roemmich, John Gilson, and Bruce Cornuelle. The main points were:

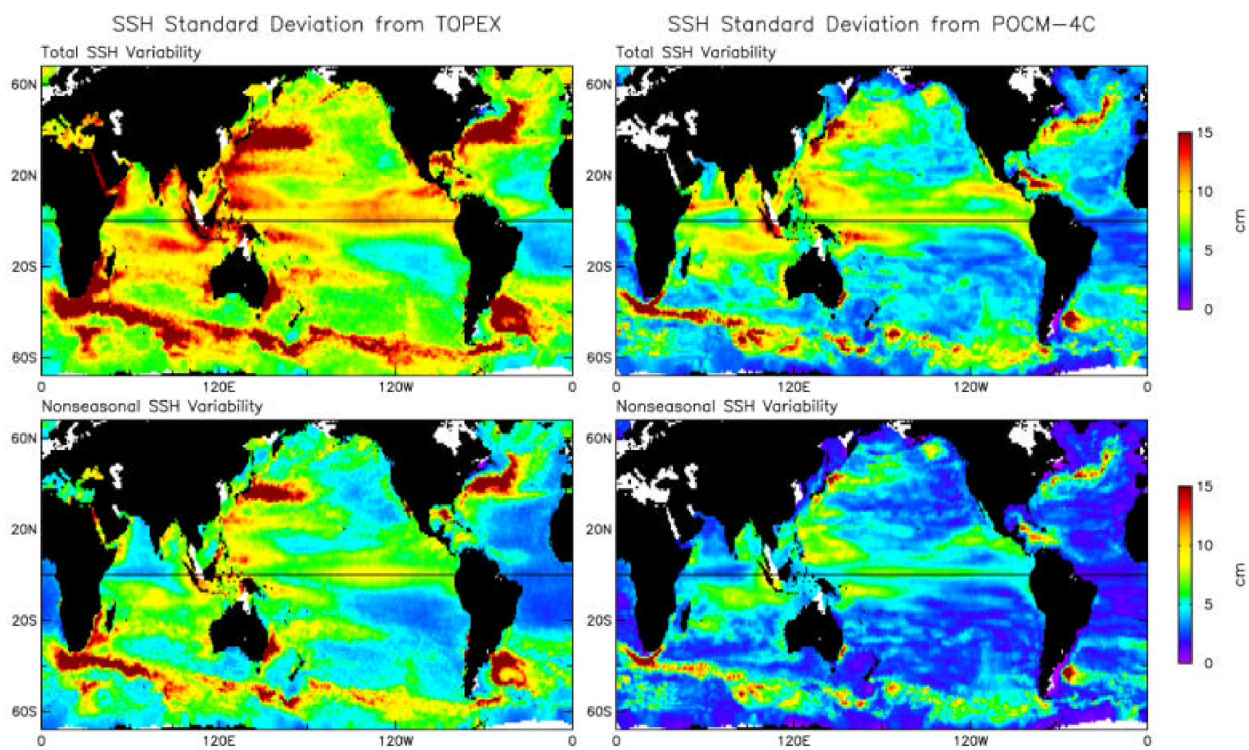
- Summary of the existing and planned observing system, and what it tells us about mean structure and variability, with an eye towards challenges to models. For example, the HRXBT section (PX37) that crosses the Kuroshio South and East of Taiwan shows small-scale but long-term features, which have significant effects on transport. These may be aliased due to the infrequent (quarterly) sampling of the HRXBT network, but similar jets show up in other HRXBT sections, such as PX06, which has a 16 year mean showing filaments in eastward-flowing currents from the Tasman Sea. The HRXBT network has also shown the importance of eddies for heat transport across section PX37.
- Data limitations mainly stem from inadequate space-time sampling. HRXBT sections are sampled at very high resolution near topography and in regions of small scales, so as to resolve the transport and heat flux, they sample only along a few tracks, and at a rate of 4/year. It is hoped that models help combine data and fill the gaps, but the error bars used on the data depend strongly on the ability of the model to accurately represent the operative physics in the region. At minimum, the HRXBT network allows good estimates of 2-d covariances in a number of representative regions.
- What measures are best to use as benchmarks or metrics for demonstrating model skill? Heat flux may be too difficult, but it’s a very physical diagnostic. Low-resolution models may not be able to get it correct, though.
- Assimilation is the best and final QC check for observations, and should be fed back to the observers for comment.
- Array Design; a list of holes in the existing network, particularly shallow-water transports in western boundary regions. These may suggest the use of gliders.

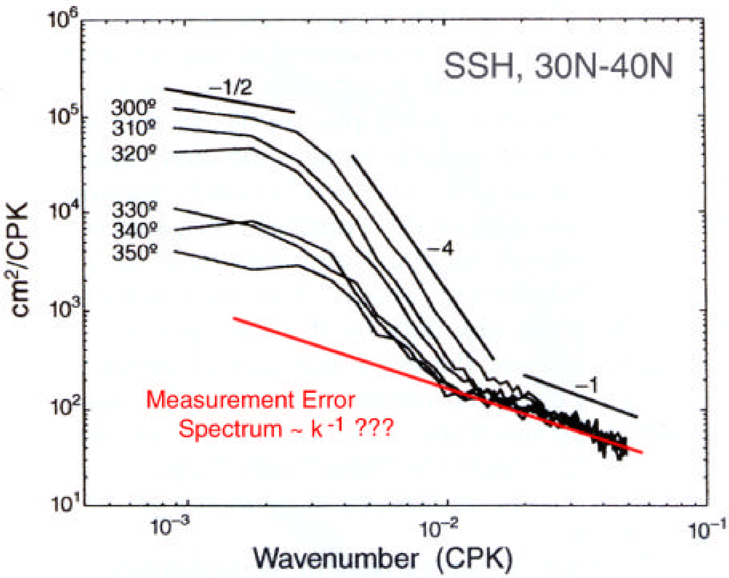
Satellite Observations Relevant to GODAE

Dudley Chelton
Oregon State University

Overview

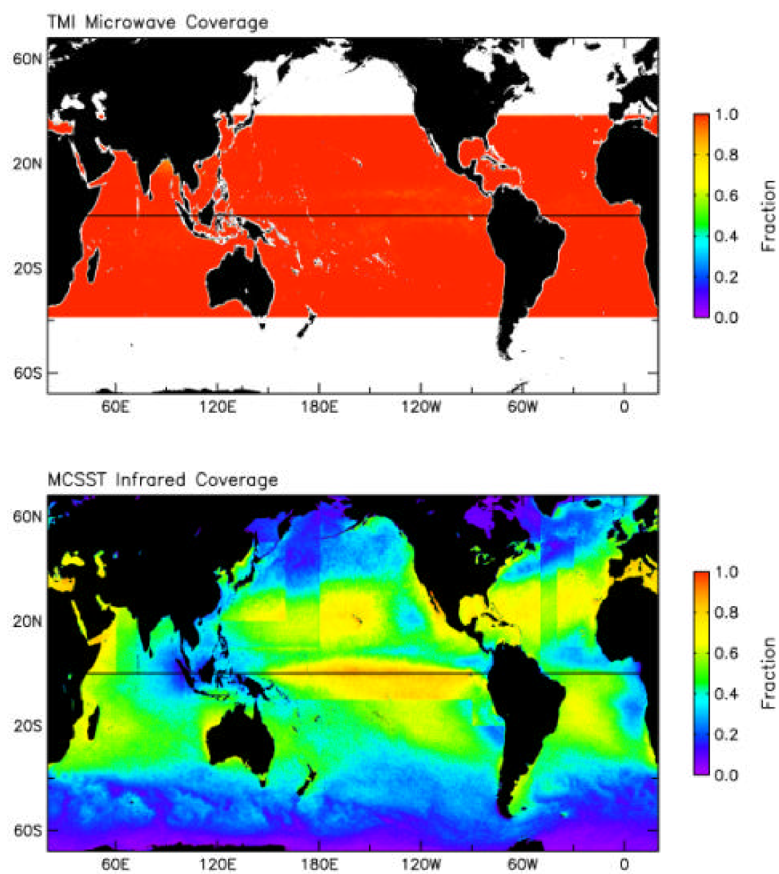
- Summary of satellite datasets useful for global ocean data assimilation
  - Altimetry for measurements of sea surface height
  - Passive microwave radiometry for measurements of SST
  - Scatterometry for measurements of surface wind stress
- Examples showing deficiencies of purely dynamical models
- Examples of new understanding of the ocean and air-sea interaction obtained from satellite data alone
- Examples showing limitations of data alone
  - There are strong needs for combining dynamics and observations through data assimilation to learn more than can be learned from either approach along
- Future outlook for satellite datasets

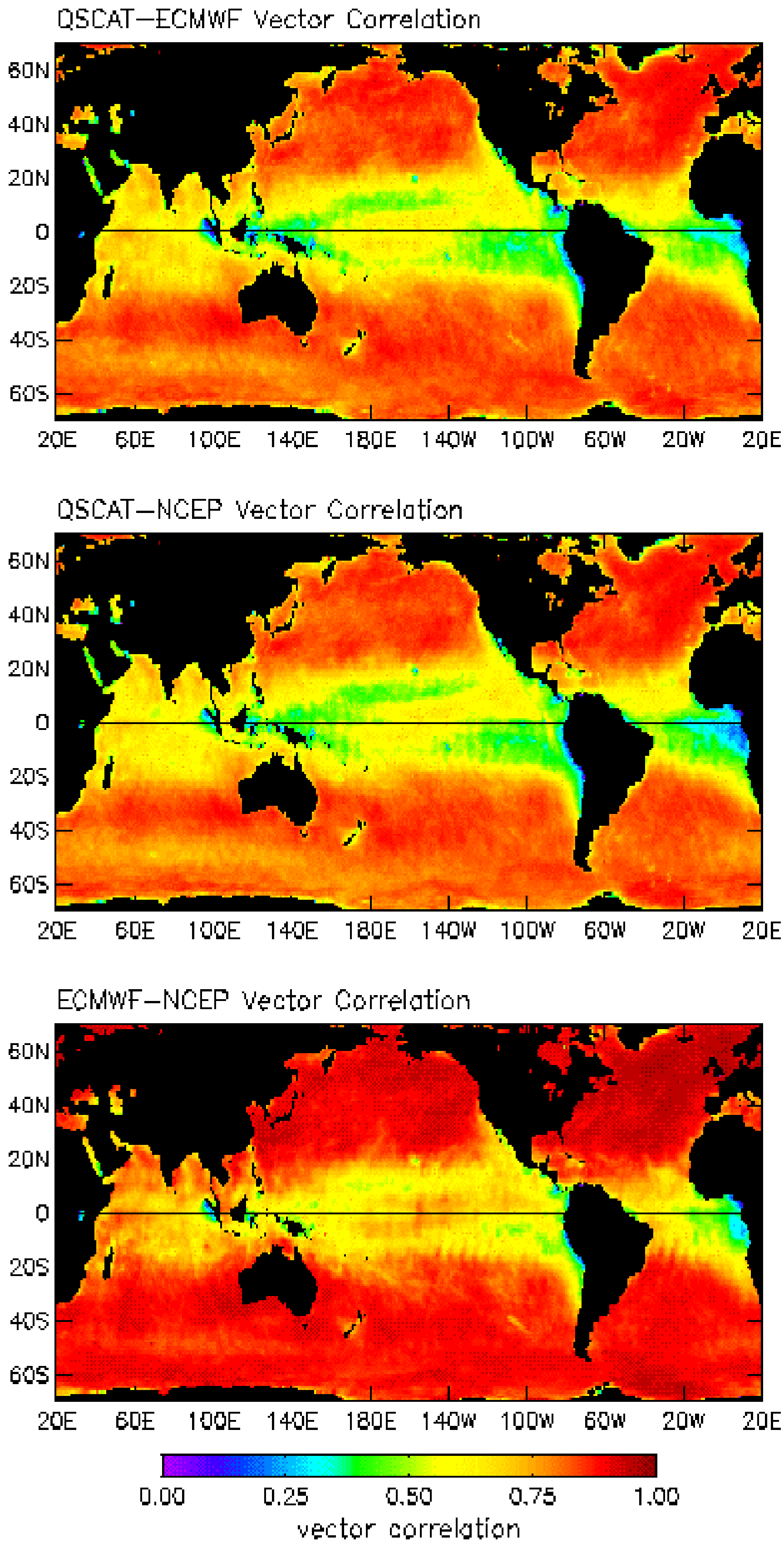


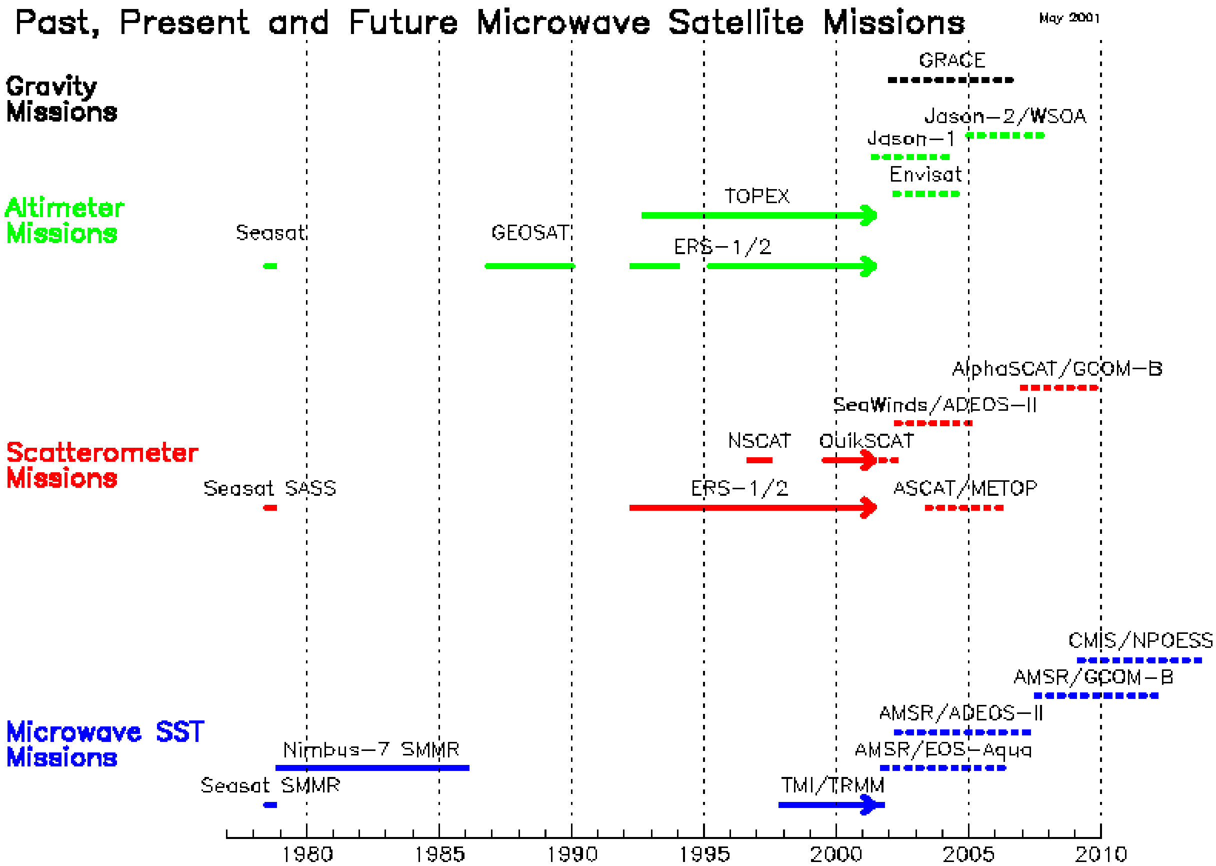


Along-track wavenumber spectra of SSH measured by TOPEX, from Stammer (1997). The redness of the measurement error spectra implies that the error covariance matrix is nondiagonal.

Weekly Average Sampling of SST




QSCAT—ECMWF Vector Correlation
60N
40N
20N
0
20S
40S
60S
20E
60E
100E
140E
180E
140W
100W
60W
20W
20E
QSCAT—NCEP Vector Correlation
ECMWF—NCEP Vector Correlation
0.00
0.25
0.50
0.75
1.00
vector correlation



Parting comments

The ocean data assimilation community would be remiss to assume that the availability of satellite data is assured for the future

- Don't be a "passive user" of satellite data

The situation is especially worrisome after ~2010

- NASA, NOAA and the U.S. Navy satellite programs need active support from the ocean data assimilation community.

Along-track and Gridded Sea Surface Products

Gregg Jacobs
NRL/Stennis

Altimeter Data Flow
Real Time Altimeter Data Availability:

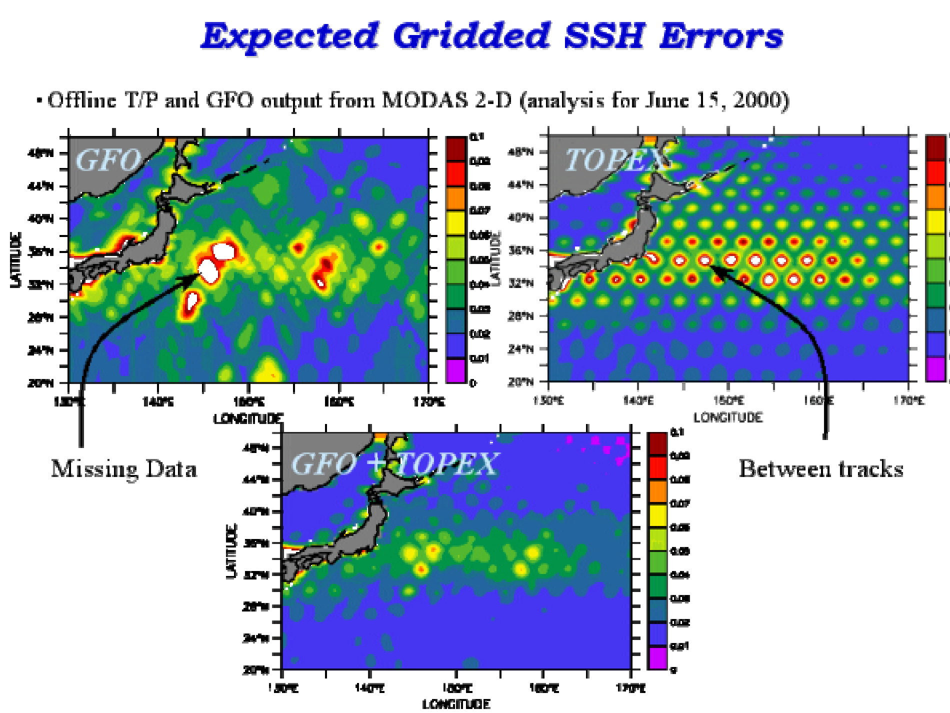
- TOPEX/POSEIDON (through JPL)
- ERS-2 (through ESA and NOAA)
- Geosat Follow-On (direct to NAVO)

2 versions of GFO and T/P of different quality (orbit solutions and ionospheric corrections)

Altimeter data streams are centralized at NAVO in the Altimeter Data Fusion Center (ADFC)

ALPS (Altimeter Processing System) Daily Analysis

- Daily QC checks
- Several processing modules each perform different QC analyses
- Each module flags suspect data uniquely
- Suspect data are examined daily to ensure that no problems have occurred



What is the data accuracy?

- SSH anomaly difference at points where satellite ground tracks cross one another indicate measurement errors (though actual oceanographic variations will affect the difference).
- RMS cross-over differences of a satellite with itself are a measure of consistency and noise within the individual measurement systems, while cross-over differences with other satellites are a measure of consistencies between measurement systems.

RMS Cross-over differences (cm) between 55S and 55N

	1 May – 15 May	16 May – 31 May	1 June – 15 June	16 June – 30 June	1 July – 7 July
GFO/GFO	11.3	11.5	11.5	10.6	9.5
TOPEX/TOPEX	8.0	7.6	7.7	8.0	7.6
ERS-2/ERS-2	12.7	12.2	12.5	12.0	N/A
TOPEX/ERS-2	11.9	11.6	11.3	11.4	N/A
TOPEX/GFO	10.6	10.3	10.5	10.1	10.3
ERS-2/GFO	13.5	12.3	12.0	12.1	N/A

To Grid or Not to Grid…

- Which is the more appropriate to use: along-track, gridded or synthetic subsurface data?
- The answer will depend strongly on the assimilation system
- In a 4DVAR approach, it is appropriate to use them all as long as we don't violate data assumptions (Gaussian distribution of erorrs, independence of errors, …)
- Each data type has information within it, and an expected error may be assigned to it.

Conclusion

- SSH along-track, gridded and synthetic subsurface data products will continue to be available for the next decade and beyond.
- Expected errors within each of these products can be assembled.
- If used appropriately (i.e., expected errors are taken into account), derived data may help to provide additional guidance to assimilation systems (i.e., reduce forecast error variance).

Altimetry, Forcing Issues and Global Assimilation

Detlef Stammer
Physical Oceanography Research Division
Scripps Institution of Oceanography

- Models show a surprising degree of realism in their simulation of specific aspects of the time-varying circulation.
- Surface forcing fields seem to be one of the key components for improving results.
- Ocean state estimation allows estimation of both the ocean state and associated surface forcing as one joint solution.
- Present results demonstrate the ability to combine various diverse data sets through ocean state estimation.
- There is no fundamental obstacle to an elaborate in situ and remote sensing data synthesis for near real-time applications for climate purposes.
- Empirical methods such as Optimal Interpolation and "nudging" invoke heat sources and sinks within the ocean interior that would render analyses of ocean heat transport difficult.

Rigorous methods solve the estimation problem in a dynamically and statistically consistent way. They include the Kalman Filter-Smoother and the adjoint method, both of which are computationally demanding.

APPENDIX C

Workshop Agenda

April 23: Assimilation Methodologies
A.M.: Presentations

Chair: Ed Harrison

8:30 Welcome and workshop goals - workshop committee
8:45 Bob Miller - mesoscale and SI analyses and forecasts
9:30 Ichiro Fukumori - climate analyses

10:15 Break

10:45 John Derber - What we can learn for the ocean from NWP data assimilation
11:30 Dick Dee - practical aspects of error covariance modeling

12:15 Lunch

P.M.: Working group breakout sessions

1:15 Session I. Real-time mesoscale assimilation
Chair: Paola Rizzoli & Harley Hurlburt

Limitations and prospects: current Navy operational assimilation: Jim Cummings

1:15 Session II. Assimilation for climate applications
Chair: Tony Rosati & Jim Carton

Limitations and prospects: current NOAA operational assimilation: Dave Behringer

3:00 Break

4:30 Reconvene joint session

April 24: Observational issues
A.M. Presentations

Chair: Breck Owens

8:30 Paul Robbins - Eulerian in situ observations
9:00 Joe Lacasce - Lagrangian in situ observations
9:30 Bruce Cornuelle - representativeness, sampling issues

10:00 Break

10:30 Dudley Chelton - satellite observations
11:30 Gregg Jacobs - along-track vs gridded altimetry
12:00 Lunch

P.M.:
1:00 Detlef Stammer - altimetry, forcing issues & global assimilation

1:45 Working group session
Chair: Bruce Cornuelle

3:00 Break

April 25

A.M.:
8:30 Workshop summary and report preparation

10:00 Break

12:00 Adjourn

APPENDIX D

Workshop Attendees

Dave Behringer - NOAA/NCEP - wd01db@ncep.noaa.gov
Anna Borovikov - NASA/GSFC - ayb@mohawk.gsfc.nasa.gov
Jim Carton - UMD - carton@metosrv2.umd.edu
Dudley Chelton - Oregon State Univ. -chelton@oce.orst.edu
Eric Chassignet, University of Miami - echassignet@rsmas.miami.edu
Mike Chin - U. Miami/JPL - Toshio.M.Chin@jpl.nasa.gov
Bruce Cornuelle, UCSD/SIO - bcornuelle@ucsd.edu
Jim Cummings/Monterey - NRL - cummings@nrlmry.navy.mil
Dick Dee - NASA/GSFC - ddee@dao.gsfc.nasa.gov
John Derber - NOAA/NCEP - john.derber@noaa.gov
Ichiro Fukumori - JPL - if@pacific.jpl.nasa.gov
Peter Hacker - U. Hawaii/IPRC - hacker@iniki.soest.hawaii.edu
Ed Harrison - NOAA/PMEL - harrison@pmel.noaa.gov
Matthew Harrison, NOAA/GFDL - mjh@gfdl.noaa.gov
Harley Hurlburt - NRL/SSC - hurlburt@nrlssc.navy.mil
Eric Itsweire - NSF - eitsweir@nsf.gov
Gregg Jacobs - NRL/SSC - jacobs@nrlssc.navy.mil
Ming Ji - NOAA/NCEP ming.ji@noaa.gov
Christian Keppenne - NASA/GSFC - clk@janus.gsfc.nasa.gov
Joe Lacasce - WHOI - jlacasce@whoi.edu
Tony Lee - JPL/NASA - tlee@pacific.jpl.nasa.gov
Julie McClean - NPS - mcclean@nps.navy.mil
Laury Miller - NOAA/NESDIS - Laury.Miller@noaa.gov
Bob Miller - Oregon State Univ. - miller@oce.orst.edu
Breck Owens - WHOI - bowens@whoi.edu
Rolf Reichle, NASA/GSFC - reichle@janus.gsfc.nasa.gov
Michele Rienecker, NASA/GSFC - michele.rienecker@gsfc.nasa.gov
Paola Rizzoli - MIT - rizzoli@ocean.mit.edu
Paul Robbins – UCSD/SIO - probbins@ucsd.edu
Tony Rosati - NOAA/GFDL - ar@gfdl.noaa.gov
Detlef Stammer - UCSD/SIO - dstammer@ucsd.edu
Michael Tippett, IRI - tippett@iri.ldeo.columbia.edu
Alberto Troccoli - NASA/GSFC - troccoli@janus.gsfc.nasa.gov
Max Yaremchuck - U. Hawaii/IPRC - maxy@soest.hawaii.edu

APPENDIX E

U.S. Assimilation Contributions to GODAE

The U.S. contributions to GODAE are focused towards developing the next generation operational global forecast capabilities at short-term (mesoscale) and seasonal-to-decadal climate scales for U.S. operational Agencies: the U.S. Navy and NOAA (National Oceanographic and Atmospheric Administration). The goals are to develop improved assimilation methodologies to integrate diverse data streams for

- real-time Navy and NOAA operational ocean activities such as maritime safety and forecasts of the coastal environment;
- initialization of seasonal-to-decadal climate forecast models; and
- estimates of the ocean state, including historical estimates akin to the atmospheric reanalyses, in support of research investigations such as CLIVAR.
- contribute to the design of an integrated observing system for mesoscale and climate applications through the assessment of observations and surface forcing fields in the context of ocean data assimilation.

The approach is to build upon existing operational capabilities, using them as the baseline against which to measure improvements. These operational capabilities constitute the core US real-time contributions. In addition, two new research and development activities have been initiated to help bring state-of-the-art ocean state estimation to quasi-operational status. The "Estimation of the Circulation and Climate of the Ocean" (ECCO) consortium and the Hybrid Coordinate Ocean Model (HYCOM) Consortium are supported through the U.S. National Ocean Partnership Program (NOPP). Funding is provided by the National Aeronautics and Space Administration (NASA), the National Science Foundation (NSF), and the Office of Naval Research (ONR). They will provide near real time demonstration products and research products. Other existing research activities in global ocean data assimilation will also contribute to the development efforts for GODAE. Some of these are briefly summarized below, but the list is not meant to be exhaustive - these contributions will be supported through U.S. R&D funding mechanisms and are subject to the usual uncertainties in such funding.

GODAE provides the context for bringing all of these ocean data assimilation developments and applications together to accelerate improvements and their transition to the operational environment. It achieves this through the provision of ready access to quality-controlled observations, through the new focused NOPP projects, and through the collaborations envisioned to build the intellectual and experiential portions of the GODAE Common.

Observations and Data Centers

The assimilation groups will rely on existing or developing data assembly centers to collect, process, and validate real time and delayed mode data. Data sources are

- *Forcing data:* NCEP and FNMOC real time analyses and predictions; ECMWF and NCEP reanalyses; Atlas's SSMI and scatterometer surface wind analyses.

- *In situ data:* profiling floats (Argo), XBT, TAO/Pirata/Triton will be accessed at the new in situ data server set up at the U.S. Fleet Numerical Meteorology and Oceanography Center (FNMOC) in Monterey. XBTs will be acquired from the GTS. Quality control procedures for these data will be applied locally. FNMOC, in cooperation with IFREMER, will host the full Argo data set on the server and will function as an Argo Global Data Assembly Center (GDAC).
- *Altimetry:* T/P, ERS-2, Jason-1, ENVISAT delayed mode data will be obtained from the NASA Physical Oceanography DAAC at the Jet Propulsion Center (JPL). Real time data will be accessed from the French AVISO/DUACS system in CLS. Near real time data, will be accessed from the Monterey server.
- *SST:* The U.S. Navy real-time products and the Reynolds SST product will be used, as will any high resolution GODAE SST product.

Product Dissemination

The real time products will be disseminated through the Live Access Server (LAS) on the GODAE server in Monterey. The NOAA operational products from the National Centers for Environmental Prediction (NCEP) will also be available from their server. Other groups will disseminate their own products, with links to them provided through the GODAE server in Monterey.

The existing capabilities, prototype plans for the GODAE development phase, and anticipated capabilities for the GODAE operational phase are briefly summarized below.

NCEP/Climate Modeling Branch (CMB)

NCEP/CMB's goal in ocean data assimilation is to maintain and improve their system for producing ocean initial conditions for seasonal to interannual forecasts with a coupled ocean-atmospheric model.

Plans for GODAE Intensive Period (2003-2005)

<u>Model</u>
The ocean model used by CMB is MOM3 from GFDL. The CMB version includes the isoneutral scheme of Gent and McWilliams (1990) for mixing tracers and the KPP boundary layer mixing scheme of Large, McWilliams and Doney (1994). It also has an explicit free surface and uses partial bottom cells to better resolve the topography. The grid is quasi-global (no Arctic Ocean) and extends from 75°S to 65°N. The resolution is 1 degree globally with enhanced meridional resolution in the tropics. There are 40 levels in the vertical and each of the top 20 levels is 10 meters thick. CMB will transition to MOM4 when it becomes available.

<u>Assimilation method</u>
The assimilation system uses a three-dimensional variational (3DVAR) method that is an extension of the work of Derber and Rosati (1989). The background error covariances are represented by Gaussian functions in the horizontal and vertical. The background error variance is allowed to vary by grid point and is represented by a function that is proportional to the

vertical gradient of temperature or salinity. The system assimilates sea surface temperature (SST), subsurface temperature and salinity profiles, and variations in sea surface height (SSH) and updates or corrects the model temperature and salinity fields simultaneously. CMB is currently planning to supplement observed salinity data with synthetic profiles of salinity produced with the technique of Maes and Behringer (2000). CMB expects to replace its current 3DVAR assimilation with a 4DVAR system by 2003.

Assimilation products
CMB will provide averages of the state variables and forcing fields on the full model grid in the netCDF format. Initially, monthly averages for 1990 and onward will be provided. It is anticipated that weekly averages will eventually be provided. Each of these files is approximately 70MB in size. They are available on the NCEP FTP server, ftp://ftp.ncep.noaa.gov/pub/cmb.

Prototype system (2001-2002)

The prototype system for the GODAE pilot phase is the current operational configuration. It operates in near real time with products available routinely each month.

U.S. Navy Operational System at FNMOC

Plans for GODAE Intensive Period (2003-2007)

Oceanographic Forecast Model
FNMOC will make output from the global ocean model component of its coupled air-sea system available for GODAE. The current candidate model for this component is the Los Alamos Parallel Ocean Program (POP) model, which is of GFDL heritage with an explicit free surface. The model will be run in a coupled mode with the Navy Operational Global Atmospheric Prediction System (NOGAPS). The model will evolve throughout the course of GODAE, with a target configuration of 1/10° latitude-longitude resolution with 30-40 levels in the vertical. Towards the end of the GODAE intensive period, the model will also be coupled to a sea ice prediction system. POP may be replaced with another similar ocean model during the course of GODAE.

Ocean Data Assimilation Method
The initial data assimilation implementation to support POP will be a 3D multivariate optimal interpolation analysis (Ocn_MVOI). The analysis variables are temperature, salinity, geopotential, and the u, v velocity components. Simultaneous 2D analyses of sea ice concentration, SST, and sea surface height anomalies (SSHA) are also performed. The system will be executed in a sequential incremental update cycle using the POP model forecast of the analysis variables as the first guess fields. Model initialization procedures, including normal mode and digital filter, are under investigation. The Ocn_MVOI analysis system will be transitioned to a multivariate 3DVAR and eventually to a 4DVAR technique.

Surface Wave Model
The WaveWatch III model will be coupled with NOGAPS and run globally on a 1/2 degree latitude-longitude grid (or finer) with 15-degree angular resolution (or finer) for the directional spectra. Surface wind stress fields generated by NOGAPS provide the atmospheric forcing. Model outputs are directional wave spectra from which a number of parameters, including

significant wave height, sea height, swell height, peak wave period and peak wave direction, are derived.

<u>Assimilation Products</u>
Unclassified FNMOC products will be available from the GODAE server in Monterey. Products will be:

- Global 3D fields of T, S, and ocean currents
- Global 2D fields of sea ice, SST and SSHA
- Global surface wave products
- Quality controlled ocean observations in real-time.

Prototype system (2001-2002)

The prototype system for the GODAE pilot phase is the operational configuration as expected to exist at FNMOC in late 2001: the Ocn_MVOI analysis system executed without a model forecast component and the WaveWatch III ocean wave model. During the GODAE development phase, FNMOC will implement a POP/Ocn_MVOI data assimilation cycle at 1/4° resolution.

U.S. Navy Operational System at NAVOCEANO

The R&D efforts from NRL/SSC are transitioned to the operational environment of The Naval Oceanographic Office (NAVOCEANO).

Plans for GODAE Intensive Period (2003-2005)

<u>Model</u>
During the GODAE intensive phase, it is anticipated that the NAVOCEANO operational system will be a 1/32° global NLOM (Navy Layered Ocean Model developed at NRL/SSC). NCOM (Navy Coastal Ocean Model developed at NRL) is a hybrid sigma-z model that is also being evaluated for use. In addition, HYCOM, a hybrid isopycnal/sigma/z (generalized coordinate model developed from MICOM) is planned as a next generation Navy system at 1/16° resolution during the GODAE time frame and is a candidate to be run at NAVOCEANO. The HYCOM Consortium is discussed further below.

<u>Assimilation method</u>
Data assimilation techniques for the ocean models are described under NRL/Stennis and the HYCOM Consortium. The Modular Ocean Data Assimilation System (MODAS) is an OI analysis using the previous analysis as the first guess and is currently running at 1/8° globally. The SSH and SST use altimeter and AVHRR MCSST products from NAVOCEANO and covariance functions derived from several years of satellite-based and SSH observations. MODAS also computes synthetic 3D grids based on the gridded SST and SSH using climatological relationships between subsurface temperature and SST and SSH derived from the historical database. Salinity is computed from the derived temperature using local climatological relationships between T and S.

Assimilation products
Unclassified NAVOCEANO products will be available from the NAVOCEANO server:

- MODAS OI SST and SSH analyses
- upper ocean temperature
- NLOM nowcast fields
- NLOM forecast fields out to 30 days, with verification statistics based on recent forecasts and model-data comparisons
- NCOM nowcast/forecast fields for the mixed layer.

Prototype system (2001-2002)

MODAS is currently running at 1/8° globally. NAVO is running an operational first generation eddy-resolving global ocean prediction system, a 1/16° global NLOM with assimilation of SST and T/P, ERS-2, and GFO altimeter data. Data from JASON-1 and Envisat will be added when available. Assimilation of hydrographic data will be added during this timeframe. Thirty-day forecasts are made each Wednesday and 4-day forecasts daily. In 2002 the plan is to use NCOM for high vertical resolution 1/8° global prediction of the mixed layer in combination with NLOM.

U.S. Navy Research System at NRL/Stennis

The ocean modeling group at the National Research Laboratory at Stennis Space Center (NRL/SSC) aims to integrate data and models developed by NRL/SSC and others into systems for performing nowcasts and forecasts of the mesoscale, large scale, and upper ocean, and to transition these systems to operational status at NAVOCEANO and FNMOC. Some current real-time capabilities can be viewed at http://www7320.nrlssc.navy.mil/global_nlom/index.html.

Model
The current model is the Navy Layered Ocean Model (NLOM) developed at NRL. During the GODAE development phase, NRL will participate in the development of the next generation model, HYCOM, as part of the HYCOM Consortium. The Navy Coastal Ocean Model (NCOM), a hybrid σ-z model developed at NRL, is also being evaluated for use. Near-term plans are to use NCOM for high vertical resolution, 1/8° global prediction of the mixed layer in combination with NLOM.

Assimilation Method
Several data assimilation techniques have been developed at NRL, including capabilities to assimilate satellite altimetry, SST, infrared frontal locations, and the downward projection of surface information. The 1/16° global NLOM – based system assimilates altimeter data using an OI deviation SSH analysis with the model field as the first guess and mesoscale covariances calculated from T/P and ERS-2 data. A statistical inference technique updates all layers of the model based on the analyzed SSH deviations, including geostrophic updates of the velocity field outside an equatorial band. The global model is then updated to produce a nowcast using slow insertion to further reduce gravity wave generation. The model is updated daily with real-time altimetry from NAVOCEANO. Currently, T/P, ERS/2 and GFO data are being assimilated and JASON-1 and Envisat data will be added when available. In addition, model deviations from full field MODAS SST analyses are assimilated by nudging. NCOM uses nudging to assimilate

3D synthetic T and S fields generated from SSH and SST fields using the MODAS system algorithms and statistics. Hydrographic profiles are added using OI. The nudging in NCOM is weak from below the surface to the base of the mixed layer to allow greater model impact on the mixed layer. See HYCOM consortium for data assimilation plans for HYCOM.

ECCO Consortium

The ECCO Consortium is a partnership between the Scripps Institution of Oceanography (SIO), JPL, and the Massachusetts Institute of Technology (MIT). Their primary goal is to provide the best possible and dynamically consistent estimate of the ocean circulation for climate analyses:

- a global ocean state estimation for the 15+ year period 1985 to present with 1/4° resolution, globally and with embedded higher-resolution efforts;
- a 50-year data synthesis with global 1° horizontal resolution;
- near real-time estimation of high frequency barotropic motion.

The Consortium activities and progress are documented at http://ecco.ucsd.edu and http://ecco.jpl.nasa.gov/odap/html/.

Plans for GODAE Intensive Period (2003-2005)

Model

The ECCO ocean state estimation is based on the MIT GCM (Marshall et al., 1997a, b) with up to 1° spatial resolution. Full global ocean state estimation capabilities with 1/4° resolution are expected by the end of FY04.

Assimilation method

The Consortium employs two optimization efforts: a 4DVAR method exploiting the Tangent-linear and Adjoint Compiler (TAMC) of Giering and Kaminsky (1997), and a Reduced-State Kalman Filter-RTS Smoother (e.g., Fukumori et al., 1999). The adjoint model uses open boundary and initial conditions, surface forcing, and internal model uncertainties as controls. Surface stress errors have been estimated from NSCAT. Ongoing computations include a 6-year estimate of the ocean circulation from 1992 to 1997.

Assimilation products and dissemination

The operational ocean state products will be focused towards best possible estimates of climate-related heat and freshwater flux:

- a 1/4° global ocean state estimate from 1985 to the present using both the Kalman filter and the adoint method;
- near real-time estimates of the full 3-D ocean state;
- real-time estimates of the barotropic motion;
- 1° global state estimate from 1950 to the present using both the Kalman Filter and the adjoint model.

Products will be disseminated through a WWW server.

HYCOM Consortium

The HYCOM Consortium is a partnership between the University of Miami/RSMAS, NRL/SSC, NOAA/AOML, the University of Minnesota, Los Alamos National Laboratory (LANL), Planning Systems Inc., Orbital Image Corp., and the U.S. Coast Guard, and the French Service Hydrographique et Oceanographique de la Marine (SHOM). The primary goal is to develop accurate forecasting capabilities for the mesoscale eddy fields, encompassing the establishment of a global real time ocean forecast system. The Consortium activities and progress are documented at http://hycom.rsmas.miami.edu.

Plans for GODAE Intensive Period (2003-2005)

Model
HYCOM (Hybrid Coordinate Ocean Model): a generalized (hybrid) coordinate model developed from the Miami Isopycnic Coordinate Ocean Model (MICOM). The hybrid coordinate is one that is isopycnal in the open, stratified ocean, but smoothly reverts to a terrain-following coordinate in shallow coastal regions, and to z-level coordinates in the mixed layer and/or unstratified seas. During the GODAE intensive period, the HYCOM Consortium will employ a global model with 1/12° and 26 coordinate surfaces.

Assimilation method
The ongoing ocean state estimation efforts use a hierarchy of assimilation methods:
Optimal interpolation (OI) and Parameter Matrix Objective Analysis (PMOA) where the OI can employ space/time-varying correlation parameters and the Cooper-Haines technique is used to project surface information to the subsurface. An adaptive filter in which unknown parameters are estimated through the model's adjoint (Hoang et al., 1997). This filter will be used to estimate the vertical correlation coefficients of the forecast error matrix. The Markov Random Field Information Filter (MRFIF) in which the horizontal covariance functions of sea surface height (SSH) and velocities are parameterized as 2^{nd}-order spatial Markov Random Fields (Chin et al., 1999).

Assimilation products and dissemination
The HYCOM products are 3D fields of the ocean state available every 3 days, and surface fields available daily. Dissemination will be via the WWW.

Prototype products (2001-2002)

The initial product is a 1/12° simulation of the North and Equatorial Atlantic Ocean, from 28°S to 70°N, with 26 coordinate surfaces. The open boundaries are treated as closed, with buffer zones. During the GODAE pilot phase, the HYCOM products will transition to global syntheses of the historical database, including the Arctic, with a 1/12° North Atlantic basin resolution and 1.4° resolution elsewhere.

NSIPP

The NASA Seasonal-to-Interannual Prediction Project (NSIPP) at NASA/Goddard Space Flight Center is directed towards seasonal-to-interannual climate prediction using coupled ocean-atmosphere-land surface general circulation models. The primary ocean data assimilation goal is to provide the best possible ocean initialization for climate prediction. NSIPP activities and progress are documented at http://nsipp.gsfc.nasa.gov.

Plans for GODAE Intensive Period (2003-2005)

Model
NSIPP uses the Poseidon quasi-isopycnal ocean model. For the GODAE operational phase, the model will be a full-water column model with an explicit free surface. The model will be run quasi-globally (no Arctic Ocean) from Antarctica to 72°N with a buffer zone at the northern boundary. The model resolution is 1/3° meridionally and 5/8° zonally with 20 isopycnal layers.

Assimilation method
NSIPP's assimilation is based both on multivariate optimal interpolation and on an Ensemble Kalman Filter (EnKF).

Assimilation products and dissemination
Assimilation products will be global, monthly averaged estimates of T, S, ocean currents and sea surface height. Products will be disseminated through the NSIPP WWW server.

Prototype system (2001-2002)

The first prototype products are based on a simple univariate OI scheme. During the GODAE development phase, the operational system will be upgraded to the multivariate OI scheme, and tests will be undertaken with the EnKF. OI and EnKF estimates in the Pacific will be generated from the historical database from 1993 to the present.

University of Maryland Research Analyses

Ocean data assimilation activities in the Department of Meteorology at the University of Maryland are directed to the ocean climate community. They are documented at http://www.meto.umd.edu./~carton. The plan is to conduct an ocean reanalysis of the historical data base over as much of the last 100 years as possible.

Plans for GODAE Intensive Period (2003-2005)

Model
The model is MOM2 run globally (without the Arctic Ocean) at 1° resolution with 30 levels.

Assimilation method
A multivariate sequential analysis method employs inhomogeneous, anisotropic, flow-dependent background error covariances with continuous incremental analysis and bias correction methods.

Assimilation products and dissemination
Monthly estimates of u, v, T, S, P are distributed by anonymous ftp through ftp://nimbus04.umd.edu/pub/outgoing/cao.

Prototype system (2001-2002)

The current system will be used to produce a 50+ year analysis. The focus is on the bias, on spatially correlated error, and on mixed layer dynamics. The analysis procedure will transition to a smoothing algorithm.

APPENDIX F

REFERENCES

Chin, T.M., A.J. Mariano, and E.P. Chassignet: Spatial regression and multiscale approximations for sequential data assimilation in ocean models, *J. Geophys. Res.*, **104**, 7991-8014, 1999.

Cohn, S. E.: Dynamics of short-term univariate forecast error covariances. *Mon. Wea. Rev*., **121**, 3123-3149, 1993.

Davis, R.E.: Drifter observations of coastal surface currents during CODE: The statistical and dynamic views. J Geophy. Res., 90, 4756-4772, 1985.

Davis, R.E.: Observing the general circulation with floats, *Deep-Sea Res.*, **38**, Suppl 1, S531-S571, 1991.

Dee, D. P.: On-line estimation of error covariance parameters for atmospheric data assimilation, *Mon. Wea. Rev*., **123**, 1128–1145, 1995.

Dee D.P., and A.M. da Silva : Data assimilation in the presence of forecast bias, *Quart. J. Roy. Meteorol. Soc.,* **124**, 269-295, 1998.

Dee D. P., and R. Todling: Data assimilation in the presence of forecast bias: The GEOS moisture analysis, *Mon. Wea. Rev.*,**128**, 3268-3282, 2000.

Derber, J., and A. Rosati: A global oceanic data assimilation system, *J. Phys. Oceanogr*., **19**, 1333-1347, 1989.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res*., **99**, 10,143-10,162, 1994.

Fukumori I., R. Raghunath, L.-L. Fu and Y. Chao: Assimilation of Topex/Poseidon altimeter data into a global ocean circulation model: How good are the results? *J. Geophys. Res.,* **104**, 25647-25665, 1999.

Fillenbaum, E.R., T.N. Lee, W.E. Johns, R.J. Zantopp: Meridional heat transport variability at 26.5N in the North Atlantic, *J. Phys. Oceanogr.,* **27**, 153-174, 1997.

Gandin, L.: Objective Analysis of Meteorological Fields, Gridromet,Leningrad. English tranlation, Israel Program for Scientific Translation, Jerusalem, 1963.

Gent P., and J.C. McWilliams: Isopycnal mixing in ocean models, *J. Phys. Oceanogr.,* **20**, 150-155, 1990.

Giering R. and T. Kaminsky: Recipes for adjoint code construction, *ACM Trans. Math. Software,* **24**, 437-474, 1997.

Heemink, A.W., M. Verlaan, and A.J. Segers: Variance reduced ensemble Kalman filtering, *Mon. Wea. Rev.,* **129**, 1718-1728, 2001.

Hoang S., R. Baraille, O. Talagrand, X. Carton and P. DeMey: Adaptive filtering: Application to satellite data assimilation in oceanography, *Dyn. Oceans Atmos.,* **27**, 257-281, 1997.

LaCasce, J. and Bower, A.: Relative dispersion in the subsurface North Atlantic, *J. Mar. Res.*, **53**, 863-894, 2000.

Large, W.G., J.C. McWilliams and S.C. Doney: Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization, *Rev. Geophys.,* **32**, 363-403, 1994.

Maes, C. and D. Behringer: Using satellite-derived sea level and temperature profiles for determining the salinity variability: a new approach, *J. Geophys. Res*., **105**, 8537‹8547, 2000.

Marshall, J., C. Hill, L. Perelman and A. Adcroft: Hydrostatic, quasi-hydrostatic and non-hydrostatic ocean modeling, *J. Geophys. Res.,* **102**, 5733-5752, 1997.

Marshall, J., A. Adcroft, C. Hill, L. Perelman and C. Heisley: A finite-volume incompressible Navier-Stokes model for studies of the ocean on parallel computers, *J. Geophys. Res.*, **102**, 5753-5766, 1997.

Mitchell H.L., and P.L. Houtekamer: An adaptive ensemble Kalman filter, *Mon. Wea. Rev.*, **128**, 416-433, 2000.

Oke, P. R., J. S. Allen, R. N. Miller, G.D. Egbert and P. M. Kosro: Assimilation of surface velocity data into a primitive equation coastal ocean model, *J. Geophys. Res.*, accepted, 2002.

Owens , W. B.: A statistical description of the mean circulation and eddy variability in the Northwestern Atlantic using SOFAR floats. *Prog. in Oceanography*, **28**, 257-303, 1991.

Parrish, D. F., and J. C. Derber: The National Meteorological Center's Spectral Statistical-Interpolation Analysis System, *Mon. Wea. Rev*., **120**, 1747-1763, 1992.

Purser, R. J., and D. F. Parrish: A Bayesian technique for estimating continuosly varying statistical parameters of a variational assimilation. Preprint, Environmental Modeling Center, NOAA/NCEP, 5200 Auth Road, Camp Springs, MD 20746, 2001.

Radakovich, J. D., P. R. Houser, A. da Silva, and M. G. Bosilovich: Results from global land-surface data assimilation methods. Presented at 5th Symposium on Integrated Observing Systems, Albuquerque, NM, 14-19 January, 2001.

Riishøjgaard, L.-P.: A direct way of specifying flow-dependent background error correlations for meteorological analysis systems, *Tellus*, **50A**, 42-57, 1998.

Roemmich, D, and J. Gilson: Eddy transport of heat and thermocline waters in the North Pacific: A key to interannual/decadal climate variability*?* *J. Phys. Oceanogr*, **31**, 675-687, 2001.

Stammer, D.: Steric and wind-induced changes in Topex/Poseidon large-scale sea surface topography observations, *J. Geophys. Res*, **102**, 20,987-21,010, 1997.

Swenson, M.S. and P.P. Niiler: Statistical analysis of the surface circulation of the California Current, *J. Geophys. Res.*, **101**, 22,631-22,645, 1996.

Wahba, G., and J. Wendelberger: New mathematical methods for variational objective analysis using splines and cross validation, *Mon. Wea. Rev.*, **108**, 1122-1143, 1980.

Wijffels, S.E., J.M. Toole and R. Davis: Revisiting the South Pacific subtropical circulation: a synthesis of WOCE observations along 32 S. *J Geophys. Res.*, **106**, 19,481-19,514, 2001.

APPENDIX G

ACRONYMS

ADFC	Altimeter Data Fusion Center
AVHR R	Advanced Very High Resolution Radiometer
ECCO	Estimation of the Circulation and Climate of the Ocean
EnKF	Ensemble Kalman Filter
ERS	European Remote-Sensing Satellite
ESA	European Space Agency
GCM	General Circulation Model
GDAC	Global Data Assembly Center
GFO	Geosat Follow-On
GODAE	Global Ocean Data Assimilation Experiment
HRXBT	High Resolution XBT
HYCOM	HYbrid Coodinate Ocean Model
KPP	K-Profile Parameterization
LAS	Live Access Server
LDE	Local Dynamics Experiment
MCSST	Multi-Channel Sea Surface Temperature
MICOM	Miami Isopycnic Coordinate Ocean Model
MODAS	Modular Ocean Data Assimilation System
MOM	Modular Ocean Model
MRFIF	Markov Random Field Information Filter
MVOI	Multi-Variate Optimal Interpolation
NLOM	Navy Layered Ocean Model
NCOM	Navy Coastal Ocean Model
NOGAPS	Navy Global Atmospheric Prediction System
NOPP	National Ocean Partnership Program
NSIPP	NASA Seasonal to Interannual Prediction Project
ODA	Ocean Data Assimilation system
OLR	Outgoing Longwave Radiation
PDF	Probability Distribution Function
PMOA	Parameter Matrix Objective Analysis
POP	Parallel Ocean Program
RMS	Root Mean Square
SSH	Sea surface height
SST	Sea surface temperature
TAO	Tropical Atmosphere Ocean
OI	Optimal Interpolation
QC	Quality Control
T/P	Topex/Poseidon
VODHub	Virtual Ocean Data Hub
VOS	Volunteer Observing Ship
WWW	World Wide Web
XBT	eXpendable BathyThermograph